UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS BIOLÓGICAS
Departamento de Fisiología Animal II

AVANCES EN LA BASE GENÉTICA DE LA ENFERMEDAD CELÍACA EN POBLACIÓN ESPAÑOLA.

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Bárbara Dema Jiménez

Bajo la dirección de las doctoras

Elena Urcelay
Concepción Núñez

Madrid, 2010

ISBN: 978-84-693-5979-2

ÍNDICE

ABREVIATURAS

APC	Células presentadoras de antígeno *(antigen presenting cells)*
APOE	*Apolipoprotein E*
AR	Artritis reumatoide
ATG16L1	*ATG16 autophagy related 16-like 1 (S. cerevisiae)*
CCR1, CCR2, CCR3...	*Chemokine C-C motif receptor1, 2, 3…*
CIITA	*Class II, major histocompatibility complex, transactivator*
CLEC16A (KIAA0350)	*C-type lectin domain family 16, member A*
CTLA4	*Cytotoxic T lymphocyte antigen* 4
CU	Colitis ulcerosa
CVID	Inmunodeficiencia común variable
DIGA	Deficiencia de IgA
DLG5	*Discs, large homolog 5 (Drosophila)*
DNA	Ácido desoxirribonucleico (*Deoxyribonucleic acid*)
DSG	Dieta sin gluten
DT1	Diabetes tipo 1
EC	Enfermedad celíaca
EGB	Enfermedad de Graves-Basedow
EII	Enfermedad inflamatoria intestinal
ELISA	*Enzyme-linked immunosorbent assay*
EM	Esclerosis múltiple
EMA	Anticuerpos anti-endomisio
ESPGHAN	*European Society for Paediatric Gastroenterology, Hepatology and Nutrition*
FoxP3	*Transcription factor forkhead box P3*
GAP	*GTP-ase activator protein*
GWAS	Estudio de asociación del genoma completo *(genome wide association study)*
HLA	Antígeno Leucocitario Humano (*human leukocyte antigen)*
HWE	Equilibrio de Hardy-Weinberg (*Hardy-Weinberg equilibium*)
ICAM1	*Intercellular adhesion molecule 1*
ICOS	*Inducible T-cell co-stimulator*
IFI	Inmunofluorescencia indirecta
IFN	Interferón
Ig	Inmunoglobulina
IL-	Interleuquina
IL18R1	*Interleukin 18 receptor 1*
IL18RAP	*Interleukin 18 receptor accessory protein*

IL23R	*Interleukin 23 receptor*
IL6R	*Interleukin-6 receptor*
iNOS	Óxido nítrico sintasa inducible
IRGM	*Immunity-related GTPase family, M*
ITGA4	*Integrin alpha 4*
LB	Linfocito B
LD	Desequilibrio de ligamiento *(linkage disequilibrium)*
LFA-1	*Lymphocyte function-associated antigen 1*
LIE	Linfocitos intraepiteliales
LPP	*LIM domain containing preferred translocation partner in lipoma*
LTh	Linfocito T colaborador (*helper*)
LTreg	Linfocito T regulador
MAF	Frecuencia del alelo minoritario *(minor allele frequency)*
MHC	Complejo Principal de Histocompatibilidad (*major histocompatibility complex*)
MICA	*MHC class I polypeptide related sequence A*
MYO9B	*Myosin IXB*
NFκB	*Nuclear factor kappa-B*
NK	Células citotóxicas naturales *(natural killer)*
NKG2A,NKG2C, NKG2D	Receptores NK G2A, G2C, G2D
NKX2-3	*NK2 transcription factor related, locus 3 (Drosophila)*
NO	Óxido nítrico
NOD2	*Nucleotide-binding oligomerization domain containing 2*
NOS2A	*Nitric oxide synthase 2, inducible*
OCTN1, OCTN2	*Organic cation transporter 1,2*
OLIG3	*Oligodendrocyte transcription factor 3*
OR	*odds ratio*
PBX3	*Pre-B-cell leukemia homeobox 3*
PCR	Reacción en cadena de la polimerasa *(polymerase chain reaction)*
PCR-SSOP	*Polymerase chain reaction-sequence specific oligonucleotide probe*
PPP6C	*Protein phosphatase 6, catalytic subunit*
Ps	Psoriasis
PTPN2	*Protein tyrosine phosphatase, non-receptor type 2*
PTPN22	*Protein tyrosine phosphatase, non-receptor type 22*
REL	*v-rel reticuloendotheliosis viral oncogene homolog (avian)*

RGS1	*Regulator of G-protein signaling 1*
SERPINE2	*Serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 2*
SH2B3	*SH2B adaptor protein 3*
SLC22A4, SLC22A5	*Solute carrier family 22 A4, A5*
SLE	Lupus eritematoso sistémico (systemic lupus erythematosus)
SNP	Polimorfismo de un único nucleótido *(single nucleotide polymorphism)*
TAGAP	*T-cell activation RhoGTPase activating protein*
T-bet	*T-box expressed in T-cell*
TCR	Receptor de células T (T cell receptor)
TDT	Test de transmisión de desequilibrio *(transmission disequilibrium test)*
TG2	Transglutaminasa 2
TGF-β	*Transforming growth factor β*
TNFA	*Tumor necrosis factor α*
TNFAIP3	*Tumor necrosis factor, alpha-induced protein 3*
WTCCC	*Welcome Trust Case-Control Consortium*

ABBREVIATIONS

Abreviaturas utilizadas en los trabajos publicados que se presentan.

BACH2	*BTB and CNC homology1, Basic leucine zipper transcription factor 2*
BTLA	*B and T lymphocyte associated*
CCR4	*CC chemokine receptor 4*
CD	Enfermedad celíaca (*celiac disease*)
CI	Intervalo de confianza (*confidence interval*)
CNV	Variación en número de copias (*copy number variation*)
EAE	Encefalomielitis experimental autoinmune (*experimental autoimmune encephalomyelitis*)
E-M	Expectación-Maximización (*Expectation-Maximization)*
eQTL	*Expression quantitative trait loci*
ETS1	*v-ets erythroblastosis virus E26 oncogene homolog 1 (avian)*
FASLG	*Fas ligand (TNF superfamily, member 6)*
FDR	*False discovery rate*
GC	Control genómico (*genomic control*)
IBD	Enfermedad inflamatoria intestinal (*inflammatory bowel disease*)
IFIH1	*Interferon induced with helicase C domain 1*
INS	*Insuline*
IRF4	*Interferon regulatory factor 4*
MMEL1	*Membrane metallo-endopeptidase like 1*
mRNA	Ácido ribonucleico mensajero (*messenger ribonucleic acid*)
MS	Esclerosis Múltiple (*multiple sclerosis*)
NFKB2	*Nuclear factor of kappa Light polypeptide gene enhacer in B-cells 2 (p49/p100)*
OR	*Odds ratio*
PC	Componente principal (*principal component*)
PCA	Análisis de componentes principales (*princpal component analysis*)
PDCD1GG2	*Programmed cell death 1 ligand 2*
PLEK	*Pleckstrin*
PTPN9	*Protein phosphatase, non-receptor type 9*
QC	Control de calidad (*Quality control*)
RUNX3	*Runt-related transcription factor 3*
SOCS1	*Supressor of cytokine signaling 1*
T	Transmitido (*transmitted*)

TC	Componentes transcripcionales (*transcriptional components*)
THEMIS	*Thymocyte selection associated*
TLR7,8	*Toll like receptors 7,8*
TNIP1	*TNFAIP3 interacting protein 1*
TNFRSF14,9,8	*Tumor necrosis factor receptor superfamily member 14, 9, 8*
TNFSF 18,4	*Tumor necrosis (ligand) superfamily member 18, 4*
U	No-transmitido (*untransmitted*)
ZMIZ1	*Zinc finger, MIZ-type containing 1*

RESUMEN/SUMMARY

RESUMEN

La enfermedad celíaca (EC) es una enfermedad inflamatoria crónica intestinal que se desencadena en individuos genéticamente predispuestos al ingerir alimentos ricos en prolaminas que provienen del trigo, cebada o centeno. Se caracteriza por ser genéticamente compleja, con múltiples genes implicados, así como interacciones entre ellos y con el ambiente. El principal factor genético descrito y que afecta a más del 95% de los celíacos, se localiza en el complejo principal de histocompatibilidad (6p21) y se debe a la combinación alélica HLA-DQ2 (*HLA-DQA1*05* y *HLA-DQB1*02*) y HLA-DQ8 (*HLA-DQA1*0301* y *HLA-DQB1*0302*). Recientemente, con la llegada de los estudios de asociación por barrido genómico (GWAS, *genome wide association studies*), se han observado otras trece regiones de susceptibilidad a EC, algunas de las cuales son comunes a otras enfermedades autoinmunes. Sin embargo, todos los loci de susceptibilidad descritos explican tan sólo el 55% del riesgo genético atribuido a EC, con lo que otros genes han de estar implicados.

El objetivo de este trabajo es la búsqueda de nuevos factores genéticos que contribuyan a la susceptibilidad a padecer esta enfermedad. Para ello se llevan a cabo estudios de asociación caso-control y familiares, por ser las estrategias más utilizadas y con mayor éxito en la detección de factores genéticos implicados en enfermedades complejas.

En primer lugar, analizamos genes previamente asociados a EC pero sin confirmación posterior. Las señales de asociación descritas en los genes *ICAM1* (19p13), *SERPINE2* (2q33), *PBX3* (9q33) y *PPP6C* (9q33), localizadas en regiones de ligamiento a EC, no se detectan en nuestro estudio en población española, a pesar de tener suficiente potencia estadística. Se descarta así su posible implicación en la susceptibilidad a EC o bien la señal estaba inicialmente sobreestimada. Sin embargo, las señales en las regiones 2q12 (*IL18RAP*, *IL18R1*) y 3p21 (*CCR1*, *CCR2*, *CCR3*…) reveladas en el estudio GWA pero no replicadas en población italiana, se detectan en nuestra muestra. No obstante, en la región 2q12 se evidencia la existencia de un riesgo débil, lo que justifica la falta de asociación observada anteriormente por limitaciones de potencia estadística.

Muchas enfermedades de carácter autoinmune y/o inflamatorio incluyendo la EC, comparten factores genéticos de susceptibilidad de forma más frecuente que la esperada. Por tanto, analizamos diversos genes previamente implicados en otras enfermedades para conocer su papel en EC. La búsqueda de señales de susceptibilidad en *IL23R*, gen asociado a enfermedad inflamatoria intestinal (EII), mostró que la variante Gln381 aumenta el riesgo a padecer celiaquía. Sin embargo, los polimorfismos que modulan el riesgo a EII situados en los genes *NKX2-3*, *ATG16L1* e *IRGM,* o a diabetes tipo 1 y esclerosis múltiple en los genes *CIITA* y *CLEC16A* (*KIAA0350*), no afectan a la predisposición a padecer EC.

El estudio de *IL6* como gen candidato, evidenció la asociación del polimorfismo *-174 G/C* con la enfermedad en niñas. Por otro lado, también seleccionado por su función, se estudió el promotor del gen *NOS2A* pero con resultado negativo; aunque esto no descarta que otros factores implicados en la regulación de la transcripción del *NOS2A* aparte de los polimorfismos de la región promotora pueden tener un papel en la enfermedad.

Por último, nuestras muestras se han unido a las de otros grupos poblacionales en una colaboración internacional. El estudio conjunto ha confirmado las señales de susceptibilidad a celiaquía de las 13 regiones génicas descritas previamente (*RGS1*, *REL*, *IL18RAP*, *ITGA4*, *CTLA4/ICOS/CD28*, *CCR3*, *IL12A*, *LPP*, *IL2/IL21*, *TNFAIP3*, *TAGAP*, *SH2B3*, *PTPN2*). Además se describen 13 nuevas regiones y se sugieren otras 13 adicionales que parecen modificar también el riesgo a padecer celiaquía. La mayoría de estas regiones contienen genes con una función inmunológica.

Los avances obtenidos en este trabajo permiten ampliar el conocimiento de la base genética de la EC y proporcionan pistas importantes sobre los procesos inmunológicos que pueden estar alterados en la enfermedad.

SUMMARY

Celiac disease (CD) is a chronic inflammatory intestinal disease triggered by the ingestion of prolamins from wheat, rye and barley in genetically susceptible individuals. CD is a genetically complex disease with several genes involved but also interactions among them and with environmental factors. The main genetic factor associated (affecting over 95% of CD individuals) is located in the Major Histocompatibility Complex (6p21). This association is due to the allelic combination HLA-DQ2 (*HLA-DQA1*05* and *HLA-DQB1*02*) and HLA-DQ8 (*HLA-DQA1*0301* and *HLA-DQB1*0302*). Recently, with the arrival of genome-wide association studies (GWAS), thirteen additional regions of susceptibility have been found. Some of them are shared with other autoimmune diseases. However, the overall genetic susceptibility conferred by these loci just explains 55% of the CD genetic risk; therefore, other genes must be involved.

The aim of this work is the search of new genetic factors that could contribute to CD susceptibility. We performed case-control and familial association studies, which are the most followed and successful strategies to detect genetic susceptibility factors in complex diseases.

First, we analyzed some genes previously associated with CD but without subsequent published confirmation. The CD associated signals described in the *ICAM1* (19p13), *SERPINE2* (2q33), *PBX3* (9q33) and *PPP6C* (9q33) genes, located in CD linked regions, were not detected in our study in the Spanish population, despite having enough statistical power to detect them. Consequently, the possible implication of these signals in CD susceptibility can be ruled out or else the signal was initially overestimated. However, the signals found in the 2q12 (*IL18RAP*, *IL18R1*) and 3p21 (*CCR1*, *CCR2*, *CCR3*…) regions, shown in GWAS but not corroborated in the Italian population, were detected in our study. The association in the 2q12 region showed a weak risk, and therefore the previously reported lack of association is probably due to statistical power limitations.

Many autoimmune and/or inflammatory diseases, including CD, share some genetic susceptibility factors more frequently than expected. Therefore, we analyzed polymorphisms in several genes previously implicated in other diseases to assess their

role in CD susceptibility. The search for susceptibility signals in the *IL23R* gene, associated with inflammatory bowel disease (IBD), showed that the Gln381 variant increases CD risk. However, polymorphisms that modulate IBD risk in the *NKX2-3*, *ATG16L1* and *IRGM* genes, or type 1 diabetes and multiple sclerosis risk in the *CIITA* and *CLEC16A* (*KIAA0350*) genes, do not seem to affect CD predisposition.

The study of *IL6* as a candidate gene shows association of the polymorphism *-174 G/C* with CD in girls. In addition, the promoter of the *NOS2A* gene was studied because of its function. A negative result was observed although other factors involved in the regulation of *NOS2A*, other than the promoter polymorphisms studied, could have a role in disease susceptibility.

Finally, our samples have been used in an international collaborative study. The susceptibility signals of the 13 regions previously described (*RGS1*, *REL*, *IL18RAP*, *ITGA4*, *CTLA4/ICOS/CD28*, *CCR3*, *IL12A*, *LPP*, *IL2/IL21*, *TNFAIP3*, *TAGAP*, *SH2B3*, *PTPN2*) have been confirmed. In addition, 13 new regions are described and other 13 are suggested as possibly modifying CD risk. Most of these regions contain genes with an immune related function.

The advances obtained in this work allow expanding the knowledge on the genetic basis of CD. They also contribute to the understanding of the immunological processes that could be implicated in the disease.

INTRODUCCIÓN

1. Definición y Epidemiología

La enfermedad celíaca (EC) es un desorden inflamatorio crónico intestinal caracterizado por una atrofia de las vellosidades intestinales, que se desencadena al ingerir alimentos ricos en gluten en individuos genéticamente predispuestos.

Hasta el inicio de este siglo la prevalencia descrita de la EC no era muy elevada, debido a que el diagnóstico se basaba en la sintomatología típica asociada. Sin embargo, tras la introducción de un "screening" serológico, ha pasado a ser común en población de origen caucásico, llegando a alcanzar una prevalencia estimada de 1 por cada 100 individuos. A pesar de ello, la EC sigue estando infradiagnosticada, y por ello se habla del "iceberg" celíaco, en el que los enfermos no detectados son considerados la porción sumergida del iceberg [1]. La prevalencia es incluso más alta en los denominados grupos de riesgo (parientes de primer y segundo grado de enfermos celíacos y personas que padecen ciertas patologías) [2]. Los factores más importantes que parecen influir en la prevalencia son tanto la frecuencia de determinados alelos que codifican la molécula HLA-DQ como una dieta rica en gluten. Así esta enfermedad es menos común en americanos de origen hispano y es muy rara en África central y Asia del Este [3, 4].

Aunque puede aparecer a cualquier edad, la EC se presenta fundamentalmente en niños, detectándose un pico de incidencia a los 9-24 meses; no obstante hay otro pico en la tercera o cuarta década de vida [5]. Algunos autores relacionan la mayor incidencia en la infancia con la interrupción de la alimentación con leche materna y la introducción del gluten en la dieta, lo que favorece la aparición de la sintomatología de la enfermedad [6-8], pero aún existe mucha controversia al respecto.

Esta enfermedad es dos veces más común en mujeres que en hombres (2:1), al igual que ocurre con otras enfermedades autoinmunes [7, 9]. Sin embargo, parece que la prevalencia entre sexos se iguala, con un *ratio* 1:1, a partir de los 65 años [10].

2. Manifestaciones clínicas y Diagnóstico

2.1. Manifestaciones clínicas

La EC es clínicamente muy heterogénea, comprende tanto pacientes asintomáticos como enfermos con una sintomatología muy severa. Además, pueden aparecer síntomas intestinales o extraintestinales. Una minoría de los enfermos, principalmente niños, debuta con una clínica típica intestinal, que consiste en malabsorción, diarrea crónica y distensión abdominal [11]. Sin embargo, de forma más frecuente aparecen signos inespecíficos como anemia, retraso pondero-estatural, pérdida de peso, osteoporosis o síntomas neurológicos [12, 13].

Existen individuos con determinadas patologías que tienen una mayor probabilidad de padecer EC, considerados por tanto grupos de riesgo, como individuos con deficiencia selectiva de IgA, diabetes tipo I, síndrome de Down, dermatitis herpetiforme y tiroiditis autoinmunes, entre otras [2].

2.2. Diagnóstico

A los pacientes con sintomatología típica o a aquellos pertenecientes a grupos de riesgo, se les realizan tests de detección de marcadores serológicos altamente específicos de la enfermedad y de elevada sensibilidad. Aunque puede evaluarse la presencia de diversos anticuerpos, lo más recomendado es la valoración mediante ELISA (*Enzyme-linked immunosorbent assay*) de los niveles de IgA anti-transglutaminasa 2 tisular recombinante humana (tTG-IgA) e IgG anti-gliadina deamidada (AGA-IgG) y por IFI (inmuno fluorescencia indirecta) confirmar la positividad con la detección de anticuerpos IgA anti-endomisio (EMA-IgA), que son los más específicos. También se cuantifican los niveles de IgA sérica total para descartar falsos negativos en pacientes deficientes en esta inmunoglobulina, en cuyo caso se les miden los niveles de tTG-IgG y EMA-IgG, cuya sensibilidad, sin embargo, es menor.

El diagnóstico también se apoya en análisis genéticos. El 95% de los enfermos porta una combinación alélica específica en ciertos genes del complejo principal de histocompatibilidad: la combinación *HLA-DQA1*05/DQB1*02*, que codifica para el heterodímero HLA-DQ2. Sin embargo, aproximadamente el 25% de la población general sana presenta esta combinación alélica, con lo que el tipaje HLA es poco específico para el diagnóstico, pero es muy útil para excluir la EC en pacientes con riesgo o sospecha, con un valor predictivo negativo próximo al 100%.

En aquellos pacientes con alta sospecha de EC por las pruebas anteriores, el diagnóstico definitivo exige la realización de una biopsia del duodeno distal. Esta patología cursa con cambios histológicos de la mucosa duodenal caracterizados por una diferencia gradual en la infiltración linfocitaria así como en la atrofia de las vellosidades intestinales. Pueden observarse distintos grados en la severidad, que han sido inicialmente clasificados por Marsh [14] y posteriormente modificados por Oberhuber y colaboradores [15] y simplificados por Corazza y colaboradores [16]. El aumento de los linfocitos intraepiteliales (CD103+ LIE), aunque no es específico de esta enfermedad, puede ser indicativo de una enfermedad latente y/o potencial cuando no se detectan cambios estructurales de la mucosa en una biopsia [17].

Todas estas pruebas diagnósticas son complementarias y además son necesarias tanto para descartar como para confirmar la enfermedad, ya que no son específicas de la enfermedad y pueden encontrarse en otras patologías.

3. Clasificación

Debido al amplio espectro (que incluye ausencia) de sintomatología clínica de la EC es difícil clasificar a los pacientes celíacos. Mientras que una minoría presenta los signos típicos de la enfermedad, muchos presentan características atípicas. Además, aunque la presentación clásica y atípica son las más detectadas, las más abundantes son las formas "sumergidas" del denominado "iceberg celíaco" [1] (Figura 1). Estas formas consisten en la forma silente, la forma latente y la potencial, que se detectan principalmente mediante cribado serológico de los grupos de riesgo o por endoscopia realizada por otras patologías [18]. En la Tabla 1 se muestran las principales características de cada forma de EC.

Figura 1: Iceberg clínico celíaco.

Tabla 1: Clasificación de la EC.

	Clásica	Atípica	Silente	Latente y potencial	Refractaria
Síntomas	Malabsorción, diarrea crónica	Síntomas digestivos inespecíficos y/o síntomas extraintestinales como anemia, pérdida de peso	Ausencia de sintomatología	Ausencia o sintomatología menor	Sintomatología recurrente o persistente
Marcadores serológicos	Anticuerpos positivos	Anticuerpos positivos	Anticuerpos positivos	Anticuerpos positivos	Anticuerpos positivos
Histología	Distintos grados de severidad de atrofia de vellosidades intestinales	Distintos grados de severidad de atrofia de vellosidades intestinales	Distintos grados de severidad de atrofia de vellosidades intestinales	Normal	Atrofia de vellosidades intestinales
LIEs	Aumento	Aumento	Aumento	Aumento	Aumento
Respuesta a DSG	Mejoría de lesiones de la mucosa y de sintomatología	Mejoría de lesiones de la mucosa y de sintomatología	Mejoría porque sí presentan un grado de enfermedad muy bajo	Mejoría porque pueden desarrollar una enteropatía clásica más tarde	No responden
Malignidad	No	No	No	No	Linfoma intestinal de células T

LIEs = Linfocitos intraepiteliales, DSG= Dieta sin gluten

Por definición, inicialmente, los pacientes con serología positiva pero con una histología normal eran excluidos del diagnóstico de celiaquía. Sin embargo, ahora se sabe que muchos pacientes presentan una muy sutil atrofia de las vellosidades intestinales.

Aunque las formas silentes y latentes son las que debutan sin sintomatología aparente, pueden llegar a ser las formas más graves, puesto que pueden desarrollarse complicaciones severas que se podrían solucionar con una dieta libre de gluten, como por ejemplo alteraciones hepáticas que pueden llevar al trasplante [19] u osteoporosis que en edad adulta es difícil de revertir [20]. A su vez pueden desarrollarse linfomas intestinales u otros cánceres, como el linfoma de células T no-Hodgkin, muy asociado a celiaquía [21]. Los pacientes adultos además, presentan otras complicaciones como la denominada EC refractaria, que aparece en un 5% de los enfermos celíacos y se caracteriza por no responder a una dieta libre de gluten, llegando a necesitar inmunosupresores para su tratamiento. Muchos desarrollan una enteropatía asociada al linfoma de células T, que presenta peor pronóstico y aumenta por tanto la tasa de mortalidad de estos pacientes [22, 23].

Los mecanismos que causan la severidad de la presentación clínica se desconocen, observándose una ausencia de correlación entre el grado de atrofia de las vellosidades intestinales y la severidad de la clínica [24].

Toda la sintomatología detectada a tiempo desaparece al eliminar de la dieta el gluten, único tratamiento actualmente utilizado a pesar de la búsqueda de tratamientos alternativos que se está llevando a cabo.

4. Inmuno-patogénesis

Muchos de los mecanismos moleculares que llevan a la respuesta inmunológica y al daño tisular del epitelio intestinal que conducen a la EC se desconocen. Existen distintos procesos de inmunidad innata y adaptativa implicados, y así componentes humorales y celulares interaccionan entre sí para dar lugar al proceso patogénico. Algunos de estos mecanismos se muestran en la Figura 2.

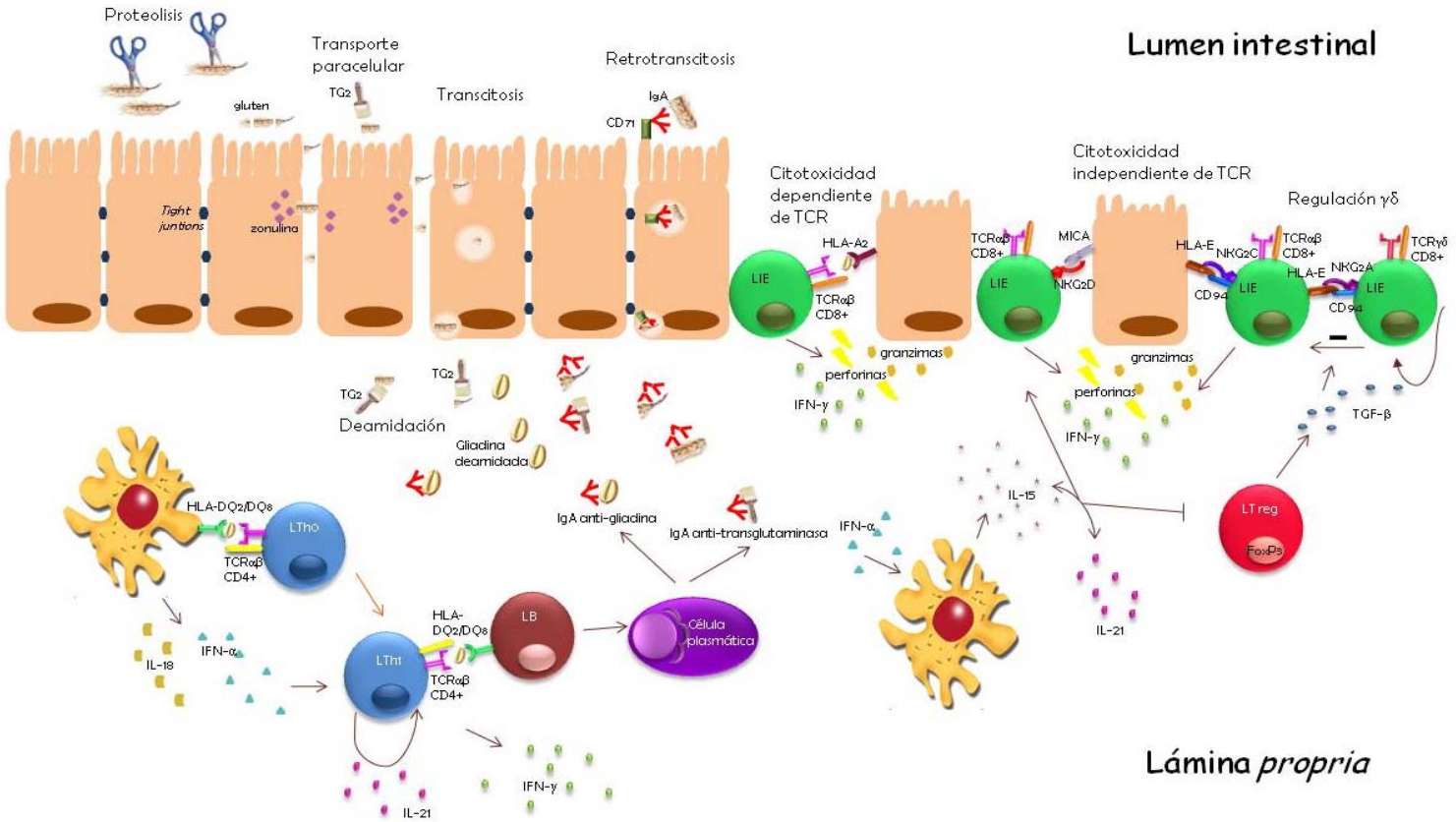

Figura 2: Inmunopatogenia de la EC.
IgA, inmunoglobulina A; TG2, transglutaminasa 2; LTh0, linfocitos T *helper* naïve; LTh1, linfocitos T *helper* 1; LB, linfocitos B; LTreg, linfocito T regulador; IFN, interferón; IL, interleuquina; LIE, linfocitos intraepiteliales; TCR, receptor de células T; MICA, *MHC class I polypeptide-related sequence A*; NKG2D, NKG2A,NKG2C, receptores NK; TGF-β, *transforming growth factor β*; FoxP3, *transcription factor forkhead box P3*.

Se han postulado distintas vías de entrada del gluten atravesando el epitelio intestinal, sin embargo muchas de estas vías aún no se han demostrado. La alteración de la permeabilidad de la barrera intestinal parece un paso inicial que permite el transporte paracelular de los péptidos de gliadina, junto con otros antígenos, a la lámina *propria* del intestino delgado. Esto puede verse favorecido, entre otros factores, al liberarse grandes cantidades de zonulina (implicada en el desensamblaje de las *tight junctions*) cuando los péptidos de gliadina se unen al epitelio intestinal [25, 26]. Además se ha detectado un transporte endocítico (transcitosis) dependiente de interferón (IFN)-γ del péptido 33-mer inmunodominante [27], aunque otros péptidos de gliadina alcanzan intactos la región basal del epitelio, uniéndose a la IgA (inmunoglobulina A) polimérica secretada y transportándose acoplados al receptor de IgA, CD71 (también conocido como receptor de transferrina) [28]. Los péptidos de gliadina pueden entonces desencadenar un proceso inflamatorio, induciendo, por un lado, una respuesta inmune adaptativa en la que se ven implicados los linfocitos T CD4+ restringidos por HLA-DQ2/DQ8, pero también pueden provocar una respuesta innata favoreciendo la producción de interleuquina (IL-)15 [29].

Los linfocitos T CD4+ específicos de gliadina, que son abundantes en la mucosa intestinal de pacientes celíacos activos, expresan el factor de transcripción que define la polarización Th1 (T *helper* 1), T-bet (*T-box expressed in T cell*) [30], e inducen la liberación de una citoquina típica de esta respuesta, IFN-γ [31]. Otra citoquina importante en la polarización Th1 es la IL-12, pero parece que no se ve alterada en pacientes celíacos no tratados [32, 33]. Sin embargo las células dendríticas de la mucosa de pacientes celíacos activos presentan niveles altos de IFN-α e IL-18, dos citoquinas que favorecen la diferenciación a Th1 estimulando la producción de IFN-γ [34, 35]. A su vez, se ha detectado recientemente que la IL-21 (producida por T CD4+) se encuentra sobreexpresada en la mucosa intestinal de pacientes celíacos activos y parece intervenir en la amplificación y estabilización de este linaje, estimulando la producción de T-bet e IFN-γ [36].

Los linfocitos T CD4+ específicos de gluten pueden estimular a los linfocitos B a producir anticuerpos IgA anti-gliadina e IgA anti-TG2, característicos de la enfermedad, pero aún se desconoce la función de estos anticuerpos en la inmunopatogenia de la misma. Los anticuerpos anti-TG2 bloquean la actividad del

enzima y se ha postulado que pueden aumentar el daño inflamatorio local debido a la implicación del enzima en reparación tisular [37].

Los linfocitos intraepiteliales juegan un papel muy importante en EC, típicamente caracterizada por un incremento en número de estos linfocitos (70-80 LIEs/100 células epiteliales en EC, <30 LIEs/100 células epiteliales en controles) [38]. Se ha sugerido que los linfocitos TCRαβ CD8+ reconocen moléculas HLA tipo I (HLA-A2) específicas de gliadina que se encuentran expuestas en las células epiteliales intestinales [39, 40]. Entonces estos linfocitos (TCRαβ CD8+) promueven una respuesta citolítica con producción de IFN-γ, granzimas y perforinas, que lleva al daño tisular. Sin embargo este ataque citotóxico de los linfocitos TCRαβ CD8+ puede ser independiente del TCR mediante la denominada reprogramación NK [38]. Se produce la sobreexpresión en su superficie de receptores activadores NK, como NKG2D, que se une a MICA (*MHC class I polypeptide-related sequence A*) expresado en células epiteliales intestinales [41], y NKG2C/CD94, que se une a una molécula HLA I no-clásica, HLA-E, de la superficie de células epiteliales intestinales [42, 43]. Esta expresión se ve favorecida por la estimulación de la IL-15 [44].

La IL-15 y la IL-21 son unos de los principales actores en la patología de la enfermedad, que de manera sinérgica promueven la proliferación y citotoxicidad de los linfocitos TCRαβ CD8+ [45]. Además, rompen el mecanismo de tolerancia oral del intestino, suprimiendo el efecto de los linfocitos T reguladores CD4+ CD25+ FoxP3+, que parecen verse aumentados en la mucosa de pacientes celíacos activos y latentes [46], sobre las células efectoras T CD4+ [47].

Por otro lado, parece que los linfocitos TCRγδ ejercen un efecto de mantenimiento de la homeostasis intestinal y reparador del daño tisular [48], y además expresan el receptor dimérico inhibidor de NK, NKG2A/CD94, suprimiendo la capacidad citotóxica de los TCRαβ CD8+ que expresan HLA-E, mediado por TGF-β (*transforming growth factor*-β) [49]. En pacientes con EC activa este efecto protector es menor [49], ya que podría verse superado por la masiva producción de IL-15, la cual es rápidamente inducida en macrófagos y/o células dendríticas por el fragmento de α-gliadina p31-43, fragmento que carece de inmunogenicidad [29].

5. Etiología

La EC presenta una etiología multifactorial, puesto que en su desarrollo están implicados tanto componentes genéticos como componentes ambientales.

5.1 Evidencia Genética

La EC es una enfermedad compleja caracterizada por carecer de una herencia mendeliana, de manera que múltiples factores genéticos llevan a la predisposición de dicho fenotipo.

La heredabilidad de esta enfermedad es elevada, es decir existe un fuerte componente genético que contribuye a la variación fenotípica total. Esta heredabilidad se conoce por los estudios de concordancia entre gemelos monozigóticos y de agregación familiar. La concordancia entre gemelos monozigóticos, parámetro que explica la contribución del componente genético a la enfermedad (indicando una ausencia de componente ambiental su proximidad al 100%), es del 75% en EC [50]. El grado de agregación familiar, comúnmente medido como el riesgo relativo entre hermanos (*sibling relative risk*, λ_s), indica el riesgo a sufrir la enfermedad en hermanos de individuos afectados con respecto al riesgo en la población general (siendo próxima al 1 la ausencia de componente genético). En EC, el riesgo relativo entre hermanos es de 20-60 [51]. Estos datos indican el importante componente genético en la EC, mayor que en otras enfermedades complejas como esclerosis múltiple (EM) (27%, λ_s=20), diabetes tipo 1 (DT1) (13%, λ_s=15) o enfermedad de Crohn (20%, λ_s=27).

5.1.1 *Genes HLA*

Los genes *HLA* (*Human Leukocyte Antigen*) del Complejo Principal de Histocompatibilidad (*Major Histocompatibility Complex*, MHC), se encuentran localizados en una región del brazo corto del cromosoma 6 (6p21) y se extienden aproximadamente a lo largo de 4Mb (Figura 3). En general son genes altamente polimórficos, es decir, cada locus presenta gran variedad alélica, aunque hay genes,

como el que codifica la cadena α de HLA-DR, que son monomórficos. La expresión de los alelos de cada locus es codominante, entendiendo esto como la expresión en la superficie celular de los dos productos proteicos de cada variante de los cromosomas homólogos. La región 6p21 se caracteriza también por presentar un fuerte desequilibrio de ligamiento (LD, *linkage disequilibrium*). Dos alelos de dos loci se encuentran en desequilibrio cuando la frecuencia en la que aparecen de forma conjunta es mayor que la que se esperaría por azar. De esta manera la diversidad haplotípica, constituida por la combinación de los alelos, se reduce en el *HLA*.

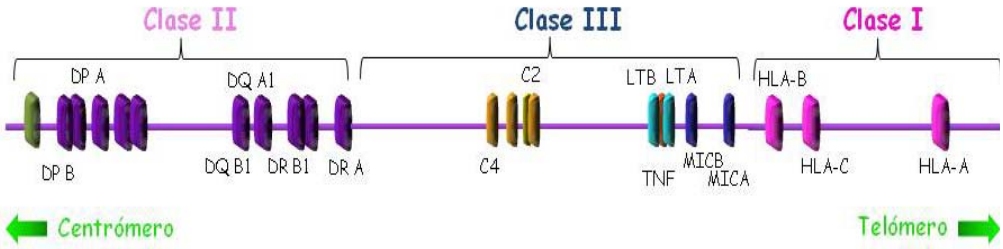

Figura 3: Localización simplificada de los genes HLA en el locus.

Los primeros estudios que mostraron asociación del *HLA* a EC, describieron la susceptibilidad aportada por el HLA-B8 (anteriormente HL-A8), detectado por técnicas serológicas [52]. Posteriormente se observó asociación con HLA-DR3 [53] y además se descubrió que se encontraba formando parte de los haplotipos ancestrales 8.1 (A1-B8-DR3), que contiene la variante HLA-B8, y 18.2 (A30-B18-DR3), ambos altamente conservados y que se asocian fuertemente a celiaquía en población caucásica [54] y de Cerdeña [55], respectivamente. Paralelamente, se detectó asociación con la variante DR7 cuando aparecía en heterocigosis con la variante DR5 [56, 57]. Estudios posteriores revelaron que la EC también se asociaba al HLA-DQ2 (anteriormente HLA-DQw2) [58], pero sobre todo a la combinación alélica *HLA-DQA1*05* y *HLA-DQB1*02* (HLA-DQ2.5), que codifican variantes de las cadenas α y β, respectivamente, de la molécula de presentación antigénica HLA-DQ. Dichas variantes alélicas pueden presentarse en *cis* (localizadas en el mismo cromosoma) o en *trans* (cada alelo en un cromosoma homólogo diferente), mostrando un desequilibrio de ligamiento muy elevado con las distintas variantes DR anteriormente citadas (Figura 4) [59, 60].

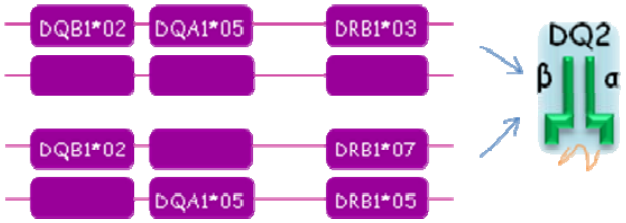

Figura 4: Combinación alélica HLA-DQ2 en *cis* o *trans*.

Aproximadamente el 90% de los enfermos celíacos presenta la molécula HLA-DQ2, pero su presencia no predice el desarrollo de la enfermedad, ya que también aparece en un cuarto de la población general sana. Por tanto, su presencia parece casi necesaria (aunque no suficiente) para el desarrollo de EC y además modifica fuertemente el riesgo a padecerla [61]. Así se puede observar un efecto de dosis génica, de modo que la presencia de dos copias del haplotipo de susceptibilidad HLA-DQ2.5, aumenta de 4 a 6 veces el riesgo de padecer EC con respecto a otras combinaciones alélicas (Tabla 2). Esto se debe principalmente a las dos copias del alelo *DQB1*02*, más que al alelo *DQA1*05*, ya que se ha observado que la combinación DQ2.5/DQ2.2 (*DQA1*05-DQB1*02*/*DQA1*02-DQB1*02*), presenta también un elevado riesgo. La presencia del haplotipo HLA-DQ2.5 *cis* en heterocigosis con otro haplotipo que no sea el HLA-DQ2.2, o el haplotipo HLA-DQ2.5 *trans*, sin embargo, confiere riesgo moderado. Este efecto de dosis génica parece que podría explicarse con el número y calidad de moléculas HLA-DQ2 localizadas en la superficie de las células presentadoras de antígeno (APC, *antigen presenting cells*) que presentan el gluten a los linfocitos T [62].

Tabla 2: Genotipos HLA-DQ2 asociados a celiaquía y efecto de dosis génica.

Tipo serológico	Genotipo DQ2	Tipo DQ	Riesgo
DR3-DQ2/DR3-DQ2	*DQA1*05-DQB1*02/ DQA1*05-DQB1*02*	DQ2.5 *cis* Homocigoto	Alto
DR3-DQ2/DR7-DQ2	*DQA1*05-DQB1*02/ DQA1*02-DQB1*02*	DQ2.5 *cis* + DQ2.2	Alto
DR3-DQ2/otra	*DQA1*05-DQB1*02*/ otra	DQ2.5 *cis* Heterocigoto	Moderado
DR5-DQ7/DR7-DQ2	*DQA1*05-DQB1*0301/ DQA1*0201-DQB1*02*	DQ2.5 *trans*	Moderado

Tabla adaptada de Dubois et al. [63]

El resto de los pacientes celíacos que carecen de la combinación alélica HLA-DQ2 (aproximadamente el 8-10%), presenta en su mayoría la molécula HLA-DQ8: *HLA-DQA1*0301* y *HLA-DQB1*0302* en *cis* [64, 65], que también ofrece distinto riesgo de susceptibilidad en función del alelo que presente en el otro cromosoma parental [66].

El grado de agregación familiar atribuido a la región del *HLA* (λ_{sHLA}) oscila de 2,2-6,6 en función de los distintos estudios [67-69], lo que indica que este factor contribuye como máximo en un 54% a la susceptibilidad de EC. Además, teniendo en cuenta que un cuarto de la población general presenta la combinación alélica HLA-DQ2.5, otros factores genéticos no-*HLA* contribuyen a la susceptibilidad a EC.

5.1.2 *Otros genes del complejo HLA*

Se ha postulado la existencia de otros factores de susceptibilidad dentro del complejo *HLA*, como el *TNFA* (*tumor necrosis factor-*α) [70-74] y *MICA* (*MHC class I polypeptide-related sequence A*) [75-77], dos genes considerados posibles candidatos por su implicación funcional, desde el punto de vista inmunológico. Sin embargo, el elevado LD de la región y la elevada potencia necesaria para efectuar estos estudios hacen muy difícil la búsqueda de otros factores de riesgo en esta región.

5.1.3 *Genes fuera del complejo HLA*

Desde la publicación de la secuencia del genoma humano en el año 2001 [78, 79], se han determinado más de diez millones de pares de bases polimórficas, cuya frecuencia puede variar dependiendo de la población considerada. Algunas de estas variaciones son neutrales, algunas son ventajosas y otras son perjudiciales para la salud [80].

En la última década, el estudio de los factores genéticos que influyen en un determinado carácter, se centró en la hipótesis "*common variant-common disease*", que propone que todas las enfermedades complejas son en gran parte debidas a variantes genéticas comunes con efectos moderados. Las variantes comunes son sinónimo de polimorfismos, definidos como variantes genéticas con una frecuencia del alelo minoritario (MAF, *minor allele frequency*) mayor del 1% [80]. Se han llevado a cabo numerosos estudios genéticos con el fin de determinar todos los factores genéticos de susceptibilidad que están implicados en una enfermedad compleja.

a. Estudios de ligamiento

Inicialmente, se llevaron a cabo mapeos genéticos para localizar genes responsables de un fenotipo, basándose en la correlación de este fenotipo con la variación genética. La forma más simple utilizada ha sido el análisis de ligamiento, que estudia la transmisión conjunta de un locus y de un determinado carácter a través de generaciones dentro de una familia con varios miembros afectados [81]. De esta manera se detectan regiones cromosómicas que tienden a compartirse entre los miembros afectados de una familia, pero que, sin embargo, difieren con respecto a los miembros no afectados. Estos estudios han tenido mucho éxito en enfermedades con una herencia mendeliana simple, sin embargo, en enfermedades complejas se encuentran numerosas limitaciones como la existencia de penetrancia incompleta, heterogeneidad genética entre familias y posible epistasis [82].

En EC se han realizado varios estudios de ligamiento, que han puesto de manifiesto más de 10 loci ligados a la enfermedad [83]. Varias de estas regiones no se han replicado, sugiriendo una heterogeneidad genética entre familias y/o la presencia de falsos positivos [84].

La región que se ha detectado más de una vez, excluyendo la región del HLA (6p21, *CELIAC1*), ha sido el locus 5q31 [85-87], considerado el locus *CELIAC2* de susceptibilidad a EC. Esta región también se ha visto ligada a la enfermedad de Crohn (*IBD5*) [88, 89], en la que los genes *SLC22A4* y *SLC22A5* localizados en esta región, y que codifican dos transportadores de cationes orgánicos (OCTN1 y OCTN2), se han descrito como los posibles genes que confieren el riesgo [90]. Sin embargo, el alto LD en esa región hace muy difícil saber con exactitud el gen/es responsables.

Otra región que inicialmente en población holandesa mostró un ligamiento significativo a EC, es 19p13 [91, 92], considerándose la región *CELIAC4*. Posteriormente, se detectó en esta región una asociación en la misma población con el gen *MYO9B* (*myosin IXB*) [93], pero que ha sido foco de mucha controversia puesto que no se ha replicado en otras poblaciones europeas incluyendo la nuestra [94-96]. En esta región también está localizado el gen *ICAM1* (*intercellular adhesión molecule* 1), gen cuya implicación en EC se ha descrito en población francesa [97], sin embargo, la confirmación o refutación de dicha asociación requiere nuevos estudios.

La búsqueda de genes implicados en EC en estas regiones de ligamiento es una estrategia para acotar la etiología poligénica de la misma, pero aún no se ha detectado el gen responsable en ninguna de las regiones descritas.

b. Estudios de asociación de genes candidatos

Los estudios de asociación se basan en la búsqueda de diferencias en la frecuencia de variantes genéticas entre individuos afectados e individuos no afectados de un determinado fenotipo, en poblaciones específicas. Dichos estudios han sido utilizados para analizar la implicación de distintos genes en numerosas enfermedades, bien afinando la localización del gen responsable en regiones identificadas por ligamiento, o bien analizando genes que por su función son posibles candidatos [98]. Estos estudios de asociación se han llevado a cabo siguiendo dos estrategias que *a priori* son difíciles de separar: una directa y otra indirecta (Figura 5). La forma directa consiste en catalogar variantes funcionalmente implicadas en regiones reguladoras o codificantes de genes, con la esperanza de que este repertorio contenga cambios que confieren susceptibilidad a la enfermedad. La forma indirecta consiste en estudiar polimorfismos que pueden ser neutrales (que no contribuyen a una variación fenotípica) pero se encuentren en LD con la variante genética que confiere el riesgo[99].

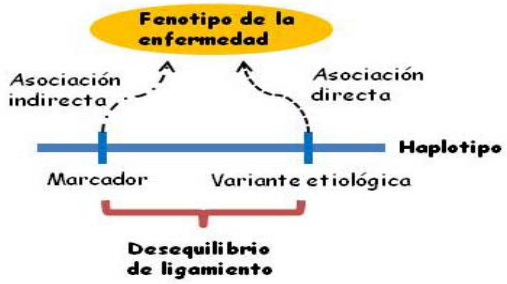

Figura 5: Estudios de asociación. (Figura adaptada de Balding et al. [100])

La forma indirecta es la estrategia más utilizada, sobre todo desde la constitución del proyecto HapMap [101, 102], que mediante el análisis inicial de polimorfismos de un único nucleótido (SNPs, *single nucleotide polymorphisms*),

cataloga los patrones haplotípicos basándose en que el genoma humano se divide en bloques de alto desequilibrio de ligamiento (elevado D') o baja diversidad haplotípica, separados por puntos de elevada recombinación (*hotspots*) que rompen el desequilibrio [103]. De esta manera, se puede establecer el mínimo grupo de marcadores necesario para recoger la práctica totalidad de variación genética en una región determinada, lo que se denomina *tagging* [104]. Con la realización de un *tagging*, se podría detectar asociación, si esta existiese con cualquiera de las variantes presentes en esta región. Sin embargo, esta situación es más compleja porque cada región genómica no tiene la misma estructura tipo bloque y porque existen distintas maneras de definir los bloques haplotípicos [105].

En los estudios de asociación se utilizan principalmente dos modelos de análisis: modelos caso-control, que utiliza individuos no emparentados; y modelos familiares, como es el TDT (*transmission disequilibrium test*) [106]. El modelo caso-control es la estrategia más aplicada para detectar asociación genética. Su objetivo es comparar la exposición a factores de riesgo entre individuos afectados y controles no afectados, ambos seleccionados de la misma población. El modelo familiar TDT consiste en la comparación entre la frecuencia con la que los alelos de un locus son transmitidos por padres heterocigotos a su hijo afectado, con la transmisión esperada típicamente mendeliana. Mientras que en el modelo caso-control se alcanza una mayor potencia, a igual número de individuos estudiados, para detectar loci de susceptibilidad, el modelo familiar implica ligamiento y asociación del locus que muestre significación, además de resolverse la presencia de falsos positivos por estratificación de la población, problema que puede afectar a los estudios caso-control.

En EC se han llevado a cabo numerosos estudios de asociación pero sin resultados concluyentes. No se han conseguido replicar muchos estudios positivos, bien por la baja potencia de los estudios que intentan replicarlo, o bien porque los resultados positivos sean consecuencia de asociaciones espurias. Los falsos positivos podrían deberse a la estratificación de la población, caracterizada por la presencia de varios subgrupos poblacionales, cada uno con frecuencias alélicas diferentes [98]. En la última década, la utilización del Control Genómico [107] (análisis de marcadores anónimos cuya frecuencia difiere entre poblaciones) para corregir la heterogeneidad poblacional está siendo de mucha utilidad, aunque, como se citó anteriormente, el

estudio familiar es otro método ampliamente usado que previene estos falsos positivos.

Uno de los locus cuya implicación en EC se ha estudiado mucho, es *CD28-CTLA4-ICOS*, localizado en la región 2q33. Inicialmente, se detectó asociación a EC en un estudio caso-control [108] y posteriormente se detectó ligamiento en la región (*CELIAC3*) [109]. A su vez esta región también se ha visto asociada a otras enfermedades: DT1 [110], enfermedad autoinmune tiroidea [111], deficiencia de IgA (DIGA) y síndrome de inmunodeficiencia variable común (CVID) [112] y artritis reumatoide (AR) [113]. Sin embargo, la variante etiológica aún se desconoce.

Como se ha citado anteriormente, esta estrategia es muy útil para el estudio de genes candidatos, es decir, aquellos que codifican productos implicados en distintos procesos de la inmunopatogénesis de EC; o aquellos relevantes en la respuestas inmunes en general y cuya implicación en otra enfermedad autoinmune pueda hacer sospechar un papel en EC. En este sentido, algunos genes interesantes serían: *IL6*, que codifica una citoquina proinflamatoria y en los últimos años ha sido considerada un importante modulador de las funciones efectoras de las células TCD4+ [114], viéndose también implicada en la diferenciación a células Th17 de forma conjunta con otras citoquinas [115]; *CIITA*, gen que codifica un importante regulador de la expresión de los genes MHC de tipo II [116]; y *NOS2A*, que codifica el enzima óxido nítrico sintasa inducible (iNOS) responsable del aumento de la síntesis del óxido nítrico (NO) en diversos tejidos y en el tracto intestinal durante procesos inflamatorios [117]. Además, estos tres genes se han visto asociados a otras enfermedades autoinmunes e inflamatorias [118-120].

La utilización de perfiles de expresión de productos génicos con el fin de identificar aquellos genes cuya expresión se ve aumentada o disminuida en el tejido correspondiente ha sido otra estrategia utilizada para la selección de genes candidatos. En EC se han llevado a cabo pocos estudios con este objetivo, aunque recientemente combinando esta estrategia con las regiones genómicas previamente ligadas a la EC, se ha descrito, en población española, la implicación de tres nuevos genes: *SERPINE2* (*serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 2)*, *PBX3* (*pre-B-cell leukemia homeobox 3*) y *PPP6C* (*protein phosphatase*

6, catalytic subunit) [121]. Sin embargo, este resultado no se ha replicado en población británica [122].

c. Estudios de asociación por barrido genómico

Los estudios de asociación por barrido genómico o GWAS (*genome wide association studies*) son otro modelo de mapeo genético que ha llevado recientemente a la identificación de decenas de loci en el genoma humano que están asociados con la susceptibilidad a padecer numerosas enfermedades complejas [123, 124]. Los GWAS son modelos carentes de hipótesis previa que buscan la asociación entre la variación genética común y un determinado carácter. Utilizan una gran selección de marcadores genéticos (10^5-10^6 marcadores) repartidos a lo largo de todo el genoma, que capturan una elevada variabilidad genética por LD, teniendo en cuenta la distribución en bloques de desequilibrio del genoma mencionada anteriormente. Los marcadores genéticos utilizados principalmente son los SNPs, debido a su alta frecuencia en el genoma humano, su baja tasa de mutación y su facilidad de automatización [80]. Analizar tal cantidad de polimorfismos puede llevar a la aparición de muchos resultados falsos positivos y por tanto es necesario imponer un factor de corrección muy exigente, así como estudios posteriores de replicación [125].

El primer estudio de barrido genómico en EC fue llevado a cabo por Van Heel y colaboradores en población inglesa [126]. Detectaron varios SNPs que mostraban asociación a EC en la región 4q27, la cual previamente no se había visto ligada a la enfermedad. La región comprende un bloque de alto desequilibrio de ligamiento que incluye los genes *KIAA1109-ADAD1-IL2-IL21* y parecen existir dos señales independientes de asociación a EC [127]. Actualmente, esta región se ha considerado un locus de susceptibilidad general de autoinmunidad porque se ha visto asociada a enfermedad de Graves-Basedow (EGB) [128], psoriasis (Ps) [129], AR [130], colitis ulcerosa (CU) [131] y DT1 [128]. Sin embargo, en el caso de DT1 parece que las variantes genéticas etiológicas son diferentes a la celiaquía y al resto de enfermedades autoinmunes [127].

Las diferencias en frecuencia entre enfermos y controles detectadas en el GWAS de celiaquía fueron más abundantes que las esperadas por azar, aunque muchas no eran significativas, probablemente debido al exigente umbral de

significación requerido por testar un número muy alto de marcadores independientes [124]. Un estudio posterior (*follow-up*) del mismo grupo [132] analizó entonces, en diferentes poblaciones europeas independientes, los marcadores que mostraban un valor de p menor de 0,003 en el estudio original de Van Heel [126] pero que no alcanzaron el umbral de significación ($p<10^{-7}$), mostrando otras siete regiones adicionales que confieren susceptibilidad a EC. Estas regiones han sido posteriormente replicadas en otras poblaciones, indicando una certera implicación de las regiones en la susceptibilidad a la enfermedad [133-135] a excepción de las regiones 2q11-2q12 (*IL18RAP*) y 3p21 (*CCR3*) que no pudieron ser replicadas en población italiana [135], sugiriendo posibles diferencias poblacionales o falta de potencia en los estudios. Estas siete regiones adicionales son:

- 1q31: contiene un bloque de desequilibrio de ligamiento de aproximadamente 70Kb que incluye el gen *RGS1* (*regulator of G-protein signaling 1*). Forma parte de la familia de proteínas reguladoras de la señalización de la proteína G, bloqueando la subunidad Gα o actuando como proteínas activadoras de la GTPasa (GAP) [136]. Este locus parece verse asociado también a DT1 [127].
- 2q11-2q12: bloque de 400Kb que comprende dos genes que codifican las subunidades del receptor de la IL-18: *IL18RAP* (*IL-18 receptor accessory protein*), *IL18R1* (*IL-18 receptor*). La IL-18 promueve la producción de IFN-γ, favorece la diferenciación a células Th1 [137] y parece que está implicada en el desarrollo de los linfocitos intraepiteliales [138]. Este locus también se ha visto asociado a DT1 (aunque con distinto efecto que en EC)[127], a enfermedad de Crohn y a CU [139].
- 3p21: es una región que contiene un *cluster* de genes de receptores de quimioquinas que incluyen: *CCR1*, *CCR2*, *CCRL2*, *CCR3*, *CCR5* y *XCR1*. Este grupo de proteínas está implicado principalmente en el reclutamiento de leucocitos [140]. Este locus se ha visto asociado a numerosas enfermedades, probablemente debido a su gran importancia en la respuesta inmunológica[141-143].
- 3q25-3q26: bloque de 70 Kb que contiene el gen *IL12A*, el cual codifica la subunidad IL12p35, que junto con la subunidad IL12p40 constituye la citoquina heterodimérica IL-12 [144]. La IL-12 media la diferenciación de células Th1 CD4+ y la producción de IFN-γ.

- 3q28: comprende un bloque de desequilibrio de 70 Kb que contiene el gen *LPP* (LIM *domain containing preferred translocation partner in lipoma*). Parece estar implicada en el mantenimiento de la forma y motilidad celular [132]. Sin embargo, poco se conoce acerca de su función. Recientemente esta región se ha visto asociada a AR aunque con un efecto opuesto al de celiaquía [145].
- 6q25: bloque de 200Kb que contiene el gen *TAGAP* (*T-cell activation Rho-GTPase activating protein*). Son proteínas Rho-GTPasas implicadas en la regulación de los cambios del citoesqueleto de actina [146]. Esta región se ha visto también asociada a DT1 [127] aunque con distinto efecto al que se detecta en EC.
- 12q24: región con señales de asociación contiguas al gen *SH2B3* (*SH2B adaptator protein 3*) que codifica la proteína SH2B3, también conocida como Lnk, involucrada en la señalización mediada por el receptor de células T, receptores de citoquinas y receptores de factores de crecimiento [147, 148]. Este locus se ha asociado a DT1 [128] y recientemente también a AR [145].

Recientemente, Garner y colaboradores [133] estudiaron dichas regiones para intentar replicar su asociación, y detectaron una nueva señal de susceptibilidad en 2q31 próximo al gen *ITGA4* (*integrin alpha 4*), aunque son necesarias replicaciones posteriores para confirmar el efecto.

La era de los GWAS ha puesto especialmente de manifiesto la existencia de loci de susceptibilidad comunes a muchas enfermedades inmunológicas, lo que sugiere un *background* genético común entre dichas enfermedades [149], posiblemente debido a mecanismos moleculares comunes. Este hallazgo llevó a Trynka y colaboradores [150] al análisis de polimorfismos que presentaban una significación baja en el GWAS de celiaquía pero que mostraban asociación a DT1, AR o enfermedad de Crohn en el estudio llevado a cabo por el WTCCC (*Wellcome Trust Case Control Consortium*) [124]. De esta manera se detectaron dos nuevas regiones asociadas a EC, 6q23 (*OLIG3-TNFAIP3, oligodendrocyte transcription factor 3/ tumor necrosis factor, alpha-induced protein 3*) y 2p16 (*REL, v-rel reticuloendotheliosis viral oncogene homolog (avian)*). Hoy se sabe que la región 6q23 muestra un efecto de susceptibilidad en AR[151, 152], DT1 [153] y lupus eritematoso sistémico (SLE) [154] y la región 2p16 se asocia a

CU [139] y a AR [155]. Ambas regiones contienen genes que codifican proteínas implicadas en la vía de activación de la familia de factores de transcripción de NFκB (*nuclear factor kappa-B*). Estos factores de transcripción controlan una amplia variedad de respuestas inmunológicas y de autoinmunidad [156].

A pesar de todos estos hallazgos, aún no se conoce el papel funcional en la enfermedad de los posibles genes candidatos presentes en estas regiones recientemente asociadas a EC, así como cuál es la variante etiológica que causa la susceptibilidad en cada región.

Todos los factores genéticos descubiertos en los GWAS, de forma conjunta, tan sólo aumentan en un 5% el riesgo genético atribuido previamente a la EC [157]. Esto indica que existen muchos factores genéticos de susceptibilidad aún no descubiertos que pueden estar implicados en la enfermedad. Un posible punto de partida sería continuar con la búsqueda de genes asociados a otras enfermedades autoinmunes y cuya función sugiera un posible papel en EC. Genes interesantes en la susceptibilidad a EC e implicados en enfermedad de Crohn serían *NKX2-3* (*NK2 transcription factor related, locus 3*), factor de transcripción que en ratones está implicado en desarrollo intestinal [158]; *ATG16L1* (*autophagy-related 16-like 1*) e *IRGM* (*immunity-related GTPase family M*), genes que codifican proteínas implicadas en autofagia [159]; e *IL23R* (*interleukin 23 receptor*), gen que codifica la subunidad específica del receptor de la IL-23, citoquina que parece intervenir en la expansión y/o estabilización del fenotipo Th17 actuando sobre los linfocitos previamente diferenciados [160]. También el estudio del gen *CLEC16A* (*KIAA0350*) podría resultar interesante tanto por su asociación a DT1 [161] y EM [162], como por estar localizado en el mismo bloque de LD que el gen *CIITA* anteriormente citado.

Por otro lado, otra estrategia que actualmente se está llevando a cabo para la búsqueda de nuevos factores genéticos de susceptibilidad, es la colaboración internacional para reunir un mayor número de muestra e intentar confirmar o refutar señales sugerentes de asociación ($10^{-4}<p>10^{-7}$) por estudios de barrido genómico previo (*follow-up* de GWAS). Esto se debe a las limitaciones en potencia estadística que se generan en los estudios GWA, a causa del elevado número de SNPs que se analizan y por tanto el exigente factor de corrección aplicado [124]. Además, puesto que

la mayoría de los GWAS evidencian efectos muy bajos en algunas de las señales reveladas (ORs=1,1-1,3), un aumento en el número de muestra será clave para poder detectar la posible asociación de variantes de bajo riesgo.

5.2. Factores ambientales

Fundamentalmente, la EC se desencadena al ingerir proteínas que proceden del trigo, cebada o centeno. Estas proteínas se denominan comúnmente gluten, aunque de forma estricta, gluten sólo correspondería a aquellas que provienen del trigo. El gluten incluye principalmente dos tipos de proteínas: gliadinas y gluteninas. Las proteínas relacionadas de la cebada y centeno que desencadenan la enfermedad son las hordeínas y secalinas, respectivamente [163].

Las gliadinas, gluteninas, hordeínas y secalinas tienen un alto contenido en residuos de prolina y glutamina, lo que les confiriere una serie de propiedades inmunogénicas. El alto contenido en prolina les proporciona resistencia frente a la digestión proteolítica de enzimas gástricas, pancreáticas y del borde en cepillo del intestino. Los residuos de glutamina son blancos de una deamidación enzimática llevada a cabo por el enzima transglutaminasa tisular o transglutaminasa 2 (TG2), enzima que se expresa constitutivamente en la lámina propia intestinal y se activa tras producirse daño tisular [164], puesto que parece que juega un papel en la reparación de tejido dañado [165]. Tras la deamidación, el péptido adquiere carga negativa (ácido glutámico), lo que favorece su unión con mayor afinidad a las moléculas de HLA-DQ2 y HLA-DQ8 [166, 167]. Los péptidos derivados del gluten se unen a la hendidura de unión antigénica que forman las cadenas α y β de las moléculas HLA-DQ2 y HLA-DQ8. Estas variantes, así como otras de HLA tipo II, en posiciones fijas de la hendidura (P4, P6 y P7 para DQ2; P1 y P9 para DQ8) reconocen determinados residuos de aminoácidos, denominados residuos "ancla", que en el caso del HLA-DQ2 y HLA-DQ8 son preferentemente aminoácidos hidrofóbicos o con carga negativa (Figura 6) [168-170].

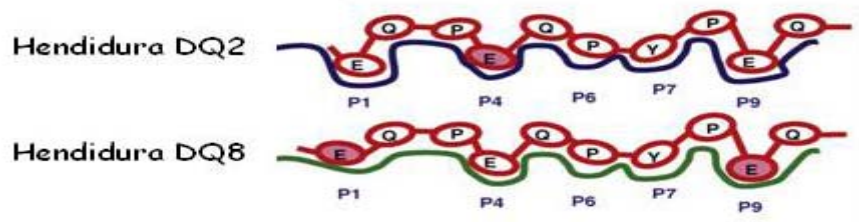

Figura 6: Epítopos de unión a la hendidura del HLA-DQ2 y DQ8. (Figura adaptada de Quiao et al. [171]).

E=ácido glutámico (coloreada cuando proviene de la deamidación), Q=glutamina, P=prolina, Y=tirosina.

Desde 1998, con la identificación del primer epítopo de gliadina reconocido por linfocitos T CD4+ restringidos por HLA-DQ2 de pacientes celíacos [167], se han identificado una gran diversidad de epítopos [171]. Mientras algunos de estos péptidos parecen ser inmunodominantes porque producen una respuesta de células T en la gran mayoría de los pacientes, otros parecen ser menos inmunogénicos. Un péptido inmunodominante en una gran parte de pacientes celíacos DQ2 positivos, es un péptido de 33 aminoácidos derivado de la α-gliadina (33-mer, del 57 al 89). El péptido 33-mer presenta seis copias de tres epítopos descritos, es resistente a la proteolisis gastrointestinal, no requiere procesamiento intracelular para su presentación y se une con muy alta afinidad al HLA-DQ2 [172].

OBJETIVOS

El componente genético que interviene en la enfermedad celíaca es muy complejo y gran parte aún se desconoce. Teniendo en cuenta esto, el principal objetivo de este trabajo ha sido la búsqueda de nuevos factores genéticos que contribuyan a la susceptibilidad a padecer esta enfermedad, siguiendo distintas aproximaciones:

1.- Estudio de genes previamente asociados a enfermedad celíaca pero sin confirmación descrita: *ICAM1*, *SERPINE2*, *PBX3*, *PPP6C, IL18RAP* y *CCR3*.

2.- Estudio de genes implicados en otras enfermedades autoinmunes que podrían tener un papel en el desarrollo de la enfermedad celíaca: *IL23R*, *ATG16L1*, *IRGM*, *NKX2-3* y *CLEC16A (KIAA0350)*.

3.- Estudio de posibles genes candidatos por su implicación funcional en la respuesta inmunitaria y/o inflamatoria: *CIITA*, *IL6* y *NOS2A*.

4.- Búsqueda de nuevas señales de susceptibilidad mediante un análisis de marcadores a gran escala utilizando una gran colección de muestras independientes.

En cada uno de los genes citados en los tres primeros puntos se han estudiado diversos polimorfismos basándonos en:

- Estudio caso-control, para determinar si existe una diferente distribución de las variantes genéticas entre pacientes con enfermedad celíaca e individuos que carecen de enfermedades autoinmunes.

- Estudio familiar, para determinar si existe una transmisión de los polimorfismos estudiados de padres heterocigotos a su hijo afectado con celiaquía diferente de la esperada por azar.

- Posible interacción de las distintas variantes genéticas analizadas con el principal factor genético de susceptibilidad HLA-DQ2.

- Posible implicación diferencial de dichas variantes entre ambos sexos.

MATERIALES Y MÉTODOS

1. Sujetos

Estudiamos un máximo de 417 pacientes con enfermedad celíaca (EC) y 956 individuos que carecen de antecedentes personales y familiares de enfermedades autoinmunes utilizados como control. A su vez, se analiza una muestra independiente de 309 enfermos con sus respectivos progenitores para llevar a cabo estudios familiares. En algunos casos se combinan ambas muestras de enfermos para alcanzar mayor potencia estadística. Todos los individuos son españoles de origen caucásico.

Las muestras de los pacientes se recogieron en el Hospital Clínico San Carlos y en el Hospital Universitario La Paz, ambos de Madrid. El diagnóstico de EC se realizó siguiendo los criterios de la Sociedad Europea de Gastroenterología, Hepatología y Nutrición (ESPGHAN) [173]. La mayoría de nuestros pacientes son pediátricos (95% son menores de 18 años), siendo la edad media de 8 años y el rango intercuartílico de 3 a 11 años. El 61% de los pacientes celíacos son mujeres; y el 2,5% presentan además deficiencia de IgA. El 92% de los celíacos presentan el factor de susceptibilidad HLA-DQ2 y de los restantes aproximadamente el 90% son portadores del factor de susceptibilidad DQ8.

Las muestras control proceden de individuos no emparentados y se obtuvieron en su mayoría de personal del Hospital Clínico San Carlos y donantes de sangre anónimos.

Todos los individuos firmaron un consentimiento informado y el estudio fue aprobado por el Comité Ético del Hospital Clínico San Carlos.

2. Aislamiento del DNA

El DNA se extrae a partir de 5 ml de sangre periférica siguiendo el método de "Salting out" [174]. Se determina la concentración de DNA por espectrofotometría (λ=260nm), y observando una pureza (A260/A280) comprendida entre 1,7-1,8, se preparan diluciones de trabajo a una concentración de DNA de 10 ng/μl.

3. Tipaje *HLA* clase II

El tipaje de los loci *HLA-DQA1* y *HLA-DQB1* se realiza mediante la técnica PCR-SSOP (*Polymerase Chain Reaction – Sequence Specific Oligonucleotide Probe*), basada en una amplificación de los loci de interés seguida de una hibridación con sondas específicas que permiten la determinación alélica [175].

4. Tipaje de polimorfismos de un único nucleótido (SNPs)

Los SNPs estudiados (Tabla 4) se seleccionan en base a la literatura previa, ya sea por su función o por resultar buenos marcadores de susceptibilidad a EC u otras enfermedades autoinmunes; o bien mediante la realización de un *tagging*. El *tagging* permite cubrir la mayor parte de la variabilidad de una región génica mediante la selección del mínimo número de SNPs, apoyándose en el desequilibrio de ligamiento entre ellos (medido como r^2). Se descartan polimorfismos que presentan un elevado r^2, es decir elevada equivalencia, con alguno de los SNPs seleccionados. Para realizarlo descargamos la región requerida (www.hapmap.org) y utilizamos el programa Haploview v4.0 (barrett2005, bioinformatics) que selecciona los SNPs dando la posibilidad de elegir la mínima frecuencia del alelo minoritario (MAF) de los polimorfismos a incluir en el *tagging*, permite incluir o excluir forzosamente determinados SNPs, así como excluir SNPs que sean equivalentes a un haplotipo formado por marcadores ya seleccionados (*aggressive tagging*). El número de polimorfismos seleccionados depende del umbral de r^2 que se elija (generalmente $\geq$80).

Tabla 4: Relación de los SNPs estudiados en este trabajo.

Región cromosómica	Gen	SNP		Localización	Cambio nucleotídico	Cambio amino-acídico
		SNP ID	Ensayo TaqMan			
1p31	*IL23R*	rs7517847	C__30369702_10	Intrón 6	T/G	-
		rs11209026	C___1272298_10	Exón 9	G/A	Arg381Gln
1q21	*IL6R*	rs8192284	*by design*	Exón 9	A/C	Asp358Ala
2q12	*IL18RAP*	rs917997	C____345197_1_	*Downstream*	G/A	-
2q33	*SERPINE2*	rs6747096	C___7614669_10	Exón 3	G/A	Asn159As
2q37	*ATG16L1*	rs2241880	C___9095577_20	Exón 8	G/A	Thr300Ala
3p21	*CCR2/CCR3*	rs6441961	C__26450554_10	Intergénico	G/A	-
5q33	*IRGM*	rs10065172	C__30593568_10	Exón	C/T	Leu105Leu
		rs4958847	C___1398968_10	3' intergénico	G/A	-
	IL12B	rs3212227	C___2084293_10	3'UTR	A/C	-
		rs6887695	C___1994992_10	5' intergénico	C/G	-
7p15	*IL6*	rs2069827	C__15860047_10	Promotor	G/T	-
		rs1800795	*by design*	Promotor	G/C	-
		rs2069840	C__15804104_10	Intrón 3	C/G	-
9q33	*PPP6C*	rs458046	C___2856723_20	Promotor	T/A	-
	PBX3	rs7040561	C__29292094_20	Intrón 2	T/A	-
10q24	*NKX2-3*	rs10883365	C__31657361_10	Promotor	G/A	-
		rs888208	C___7465608_10	3'UTR	A/G	-
16p13	*CIITA*	rs3087456	C__15793789_10	Promotor	A/G	-
		rs4774	C____381733_10	Exón 12	G/C	Gly500Ala
	KIAA0350 (CLEC16A)	rs7203459	C___2975039_10	Intrón 21	T/C	-
		rs2903692	C__15941578_10	Intrón 21	G/A	-
		rs6498169	C__29080959_10	Intrón 21	A/G	-
17q11	*NOS2A*	rs2779251	C__15932163_10	Promotor	C/T	-
		rs2779248	C___2593688_10	Promotor	A/G	-
19p13	*ICAM1*	rs281432	C____944685_10	Intrón 2	C/G	-
		rs1799969	*by design*	Exón 4	G/A	Gly241Arg
		rs1801714	C___8726336_10	Exón 5	C/T	Pro352Leu
		rs5030400	C__43870552_10	Exón 7	C/T	Arg478Trp

El tipaje de los SNPs se realiza mediante tecnología TaqMan [176], utilizando el equipo ABI PRISM 7900HT (Applied Biosystems, Foster City, CA, USA). Esta técnica se basa en una PCR a la que se le añaden sondas específicas para cada alelo del SNP que se va a estudiar. Cada sonda se encuentra marcada en su extremo 5' con el fluoróforo FAM o VIC y en su extremo 3' presenta un *quencher* con actividad "secuestradora" de fluorescencia. La amplificación del fragmento de DNA que contiene el alelo complementario al de la sonda conlleva un aumento exponencial de la fluorescencia emitida por dicha sonda cuando es degradada por la DNA polimerasa (Taq Gold) con su actividad exonucleasa 5´→3´ y por tanto se libera del *quencher*, a medida que avanza la polimerización. De este modo, dependiendo de la fluorescencia observada sabremos el alelo/s presente en la muestra (figura 10).

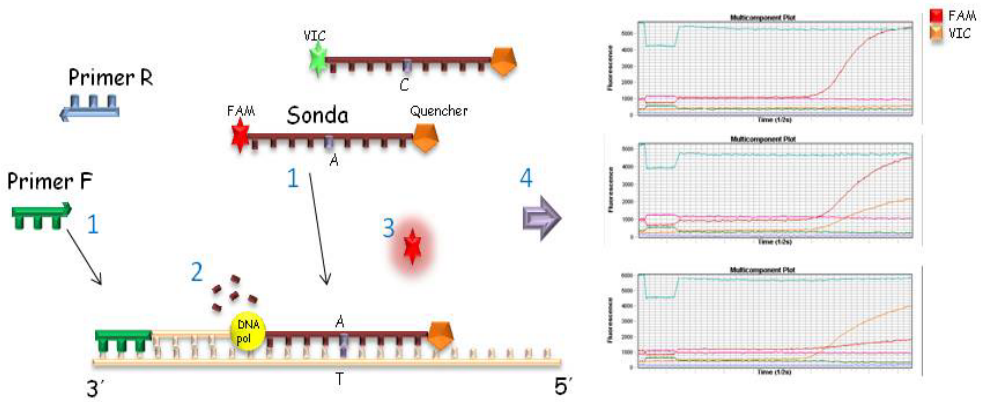

Figura 10: Representación simplificada del ensayo TaqMan. 1. Hibridación de sonda y *primers*. 2. DNA Polimerasa: polimerización 3´→5´y degradación 5´→3´. 3. Emisión de fluorescencia liberada. 4. Representación de la emisión de fluorescencia exponencial detectada.

Las condiciones de PCR generales utilizadas son:

- 2 min. a 50ºC, que supone la activación del enzima amperasa.
- 10 min. a 95ºC, activación polimerasa.
- 40 ciclos: 15 seg. a 95-93ºC (desnaturalización) y 1 min. a 60ºC (hibridación y elongación).

6. Tipaje de polimorfismos microsatélites

Para determinar el genotipo de los microsatélites se llevó a cabo una PCR con primers o cebadores específicos, uno de los cuales se encuentra marcado en su extremo 5´ con un fluoróforo (HEX o FAM).

- *NOS2A*: pentanucleótido $(CCTTT)_n$ (rs3833912).
Primer forward: 5'-Fam-ACCCCTGGAAGCCTACAACT-3'
Primer reverse: 5'-GCCACTGCACCCTAGCCTGTCTCA-3'

- *NOS2A*: inserción/deleción TAAA (4 ó 3 repeticiones, respectivamente) (rs12720460).
Primer forward: 5'-Hex-TGGTGCATGCCTGTAGTCC-3'
Primer reverse: 5'-GAGGCCTCTGAGATGTTGGTC-3'

La amplificación se lleva a cabo en las condiciones óptimas para cada microsatélite: 10 min. a 95ºC (activación de la polimerasa), 37 ciclos: 30 seg. a 95ºC (desnaturalización), 30 seg. 60ºC (hibridación), 40 seg. a 72ºC (elongación) y finalmente 7 min. a 60ºC para una elongación final. En el microsatélite inserción/deleción TAAA, la temperatura de hibridación es de 55ºC.

El producto amplificado se mezcla con formamida (agente desnaturalizante) y un marcador de tamaño (Gene Scan ROX 400HD, Applied Biosystems, Foster City, CA, USA) y se desnaturaliza (3min. a 95ºC). Posteriormente, mediante una electroforesis capilar llevada a cabo en el secuenciador ABI PRISM ® 3100 *Genetic Analyzer*, se separan los fragmentos de DNA en función de su tamaño y se detecta y analiza la fluorescencia emitida por los mismos (Figura 11).

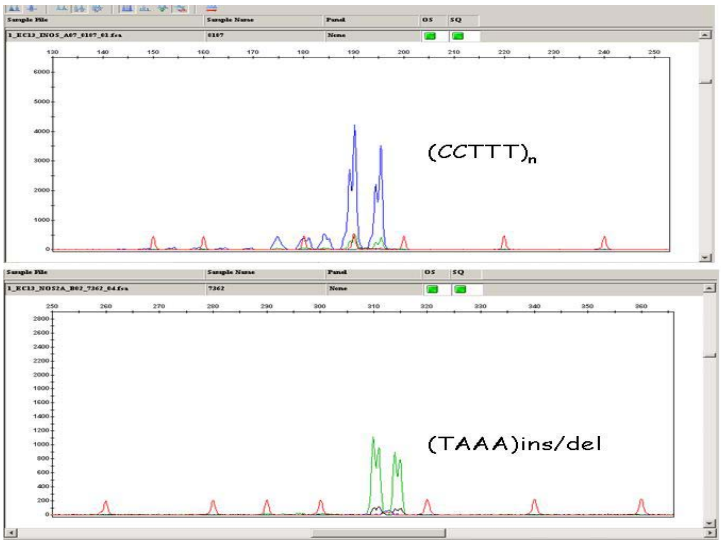

Figura 11: Representación de las áreas de fluorescencia emitida al migrar los fragmentos amplificados correspondientes a los alelos del microsatélite.

7. Análisis estadístico

Los datos obtenidos del genotipado se recopilan en una base de datos de Microsoft Access en la que ya existen otros datos como sexo y edad de los individuos, y en el caso de los pacientes, la edad de debut y otras características clínicas.

Para todos los marcadores estudiados se evalúa como primer paso que se ajusten a las proporciones esperadas bajo las condiciones del equilibrio de Hardy-Weinberg en los individuos control [177].

En el estudio caso-control se comparan las frecuencias genotípicas, alélicas y de portadores aplicando el test χ^2 o test exacto de Fisher (cuando los valores esperados son menores de 5), utilizando el programa Statcalc (EpiInfo v6).

El desequilibrio de ligamiento entre los marcadores, existente cuando dos alelos de dos loci se encuentran conjuntamente con una frecuencia mayor a la esperada por azar, fue analizado mediante el cálculo de intensidad expresado con el coeficiente D'[178] o r^2 y su significación determinada con la aplicación del test χ^2 a las tablas de contingencia 2x2 correspondientes.

En aquellas regiones génicas en las que se estudian varios marcadores en desequilibrio, las frecuencias haplotípicas son estimadas con el algoritmo EM (*Expectation-Maximization*), ofrecidas por el programa Arlequin v2.000 [179] o por el programa Haploview v4.0 [180].

El estudio familiar se llevó a cabo mediante la aplicación del TDT (*Transmission Disequilibrium Test*), en el que se determina si la frecuencia de transmisión de un alelo o haplotipo al hijo enfermo es significativamente diferente a la que cabría esperar (probabilidad de transmisión al 50%). Este test sólo utiliza datos de padres con genotipo heterocigoto [181].

En algunos casos, se analizó la posible existencia de interacción entre los polimorfismos estudiados. Este análisis se realizó comparando la distribución de las frecuencias genotípicas de cada polimorfismo, estratificado por la distribución genotípica de los otros polimorfismos con los que se quiere testar la interacción.

Para la realización de los meta-análisis se utilizó el test de Mantel-Haenszel, que combina varias odds ratio, teniendo en cuenta la homogeneidad o no entre los distintos datos (test de Breslow-Day), todo incluido en el programa Review Manager (RevMan) 5.0 (Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2008).

ESTUDIOS REALIZADOS

1.- ESTUDIO DE GENES PREVIAMENTE ASOCIADOS A CELIAQUÍA

I. *ICAM1* R241 is not associated with celiac disease in the Spanish population.

II. Lack of replication of celiac disease risk variants reported in a Spanish population using an independent Spanish sample.

III. Association of *IL18RAP* and *CCR3* with celiac disease in the Spanish population.

I. *ICAM1* R241 is not associated with celiac disease in the Spanish population

Bárbara Dema[a], Alfonso Martínez[a], Isabel Polanco[b], Carlos Maluenda[c], Miguel Fernández-Arquero[a], Emilio G. de la Concha[a], Elena Urcelay[a], Concepción Núñez[a]

[a] Servicio de Inmunología Clínica, Hospital Clínico San Carlos, 28040 Madrid, Spain
[b] Servicio de Gastroenterología Pediátrica, Hospital La Paz, 28046 Madrid, Spain
[c] Servicio de Pediatría, Hospital Clínico San Carlos, 28040 Madrid, Spain

Human Immunology (2008); 69 (10): 675-678

Resumen

La enfermedad celíaca (EC) es una enfermedad crónica inflamatoria intestinal que se desarrolla en individuos genéticamente susceptibles tras la ingesta de gluten. El gen *ICAM1*, localizado en la región ligada a EC 19p13, codifica una molécula de adhesión intracelular (ICAM-1) implicada en procesos inflamatorios. Los niveles de ICAM-1 se han visto aumentados tanto en suero como en biopsias intestinales de pacientes celíacos. Además, en población francesa se ha descrito recientemente una asociación del polimorfismo G241R del gen *ICAM1* con EC. El objetivo de este estudio ha sido analizar el papel de polimorfismos del gen *ICAM1* en la susceptibilidad a EC en población española. Llevamos a cabo un estudio caso-control con 608 pacientes con EC y 537 individuos sanos utilizados como control, y un estudio familiar incluyendo 231 tríos. Se analizaron cuatro polimorfismos de un único nucleótido (SNPs) del gen *ICAM1*: tres no-sinónimos, R478W (rs5030400), P352L (rs1801714) y G241R (rs1799969); y uno intrónico, rs281432. A pesar de presentar más del 98% de potencia estadística para detectar la asociación descrita en población francesa (odds ratio=1.7), no encontramos diferencias en frecuencias genotípicas o alélicas del polimorfismo G241R entre nuestros pacientes celíacos y controles, y tampoco observamos diferencias en los otros SNPs analizados. Por tanto, en nuestra población, nuestros resultados descartan el importante papel previamente descrito del polimorfismo *ICAM1* G241R en EC.

Abstract

Celiac disease (CD) is a chronic intestinal inflammatory disease that develops in genetically susceptible individuals after gluten ingestion. The *ICAM1* gene, located in the CD linkage region 19p13, encodes an intercellular adhesion molecule (ICAM-1) involved in inflammatory processes. Increased levels of ICAM-1 were observed in intestinal biopsies and in sera of CD patients. In addition, an association between the *ICAM1* polymorphism G241R and CD patients has been recently described in a French population. Our aim in this study was to analyze the role of *ICAM1* polymorphisms in CD susceptibility in the Spanish population.We performed a case–control study with 608 CD patients and 537 healthy control individuals and a family study including 231 trios. Four *ICAM1* single nucleotide polymorphisms (SNPs) were analyzed: three nonsynonymous, R478W (rs5030400), P352L (rs1801714) and G241R (rs1799969); and one intronic, rs281432. Despite having above 98% statistical power to detect the association described in the French population (odds ratio=1.7), we did not find any differences in genotypic or allelic frequencies of the G241R polymorphism between our CD patients and controls, and no differences were observed when the other SNPs were analyzed. Therefore, in our population our results discard the important previously described role of *ICAM1* G241R in celiac disease.

KEYWORDS: celiac disease susceptibility; intercellular adhesion molecule; ICAM-1; CD54; *CELIAC4.*

Introduction

Celiac disease (CD) is a chronic inflammatory disease caused by gluten (or related proteins) ingestion in genetically susceptible individuals. Although HLA is the main genetic factor involved in CD susceptibility, genetic components outside this complex must be also relevant. Regions resulting from linkage studies offer several genes that, according to their function, may be good candidate susceptibility genes. *CELIAC4* is located on 19p13 and no obvious gene explains the linkage described. The myosin IXB (*MYO9B*) gene emerged as the most plausible candidate after the positive association described in a case–control study performed in a Dutch population [1]. However, susceptibility caused by *MYO9B* was not replicated in additional populations [2– 4], including in our Spanish population [5], and its role as the disease-causing gene at *CELIAC4* has been questioned [6]. Curley and colleagues [7] indicated two additional genes in this region with small effects in CD risk in the Dutch population, *CYP4F3* and *CYP4F2*, but no replication confirmed this association either. An additional candidate gene on 19p13 is *ICAM1*, which encodes the immunoglobulin-like intercellular adhesion molecule ICAM-1 (CD54) that mediates interactions between cells through binding to its natural ligands, lymphocyte function associated antigen-1 (LFA-1) and Mac-1. It is expressed by several cell types, including endothelial and epithelial cells, lymphocytes, and monocytes. This expression may be upregulated by proinflammatory cytokines as interferon-γ, interleukin-1, and tumor necrosis factor-α. ICAM-1 is involved in immunological processes such as cellular extravasations and cellular migration during inflammation, as well as in T-cell activation [8]. Increased expression of ICAM-1 after gluten exposure was observed in biopsies from active CD patients when compared with treated patients (after gluten-free diet) and healthy controls [9]. Patients with active CD also demonstrated higher levels of ICAM-1 in sera [10]. However, the increased ICAM-1 levels observed in CD patients could be cause or consequence of the inflammatory process. Several polymorphisms have been described in the *ICAM1* gene, with the potentially functional G241R (rs1799969) and K469E (rs5498) the most extensively studied. In the French population, an increased frequency of R241 was observed in CD patients [11], although this association was not reported in an Indian population [12]. Those polymorphisms were also studied in several autoimmune diseases such as multiple sclerosis [13,14], type 1 diabetes [15,16],

inflammatory bowel disease [17,18], and Graves' disease [19], with variable results. We aimed to study non-synonymous or potentially functional *ICAM1* polymorphisms in another European population to determine the role of this gene in CD susceptibility.

Subjects and Methods

Subjects

We performed a case–control study including 608 white Spanish celiac disease patients and 537 ethnically matched healthy controls. A family study was also performed including both parents of 231 CD patients (included in the 608 previously analyzed). All samples were collected from two hospitals in Madrid (Hospital Clínico San Carlos and Hospital La Paz). Celiac disease was diagnosed following the European Society for Paediatric Gastroenterology and Nutrition criteria [20]. The median age for diagnosis was 8.7 years and the interquartile range was 4–12.5 years. Sixty-one percent of CD patients were female. HLA-DQ2 positivity was 92% in our CD patients. Written informed consent was obtained from all patients. Ethical approval for the study was obtained from the ethics committee of the Hospital Clínico San Carlos.

Genotyping

We analyzed four single nucleotide polymorphisms (SNPs) on the *ICAM1* gene: one intronic (rs281432) and the three available nonsynonymous polymorphisms present in Caucasian populations [rs5030400 (R478W), rs1801714 (P352L), and rs1799969 (G241R)]. All these polymorphisms were genotyped using TaqMan genotyping assays (C__944685_10, C__43870552_10, C__8726336_10, and rs1799969 assay by design, respectively) under conditions recommended by the manufacturer (Applied Biosystems, Foster City, CA). The missense mutation K469E could not be analyzed because of technical problems (the assay could not be designed by Applied Biosystems) but this SNP was not significantly associated with CD in the French population. The genotyping success rate was above 98% for all polymorphisms in CD patients and controls.

Statistical analyses

Genotypic, allelic, and haplotypic frequencies between patients and controls were compared using the χ^2 test or Fisher's exact test when expected values were below 5 (calculations performed with EpiInfo v5 software). Haplotypic frequencies were estimated using the expectation-maximization algorithm implemented using Arlequin v2.000 software, with the number of iterations set at 5,000 and initial conditions at 50, with an ε value of 10^{-7}. Family data were analyzed using the transmission disequilibrium test (TDT), which uses only information from heterozygous parents to assess the preferential transmission of one allele or haplotype over the other. Statistical power calculations were performed using EpiInfo v5 software.

Results

All polymorphisms conformed to Hardy–Weinberg expectations in the control population. The allele, genotype, and haplotype frequencies of the polymorphisms studied did not demonstrate differences between CD patients and healthy control individuals (Tables 1 and 2).

Table 1: Genotypic and allelic frequencies of *ICAM1* SNPs in CD patients and controls, and TDT results.

	CD		Controls		p	OR (95% CI))	TDT	
							T:U	p
rs281432	N=605	%	N=535	%				
CC	154	25.5	136	25.4				
CG	317	52.4	260	48.6				
GG	134	22.1	139	26.0	0.27			
C	625	51.7	532	49.7				
G	585	48.3	538	50.3	0.36	0.93 (0.78-1.09)	106:118	0.23
G241R (rs1799969)	N=608	%	N=537	%				
GG	519	85.4	443	82.5				
GR	83	13.7	91	16.9				
RR	6	1.0	3	0.6	0.22			
G (G)	1121	92.2	977	91.0				
R (A)	95	7.8	97	9.0	0.29	0.85 (0,63-1.16)	43:32	0.13
P352L (rs1801714)	N=608	%	N=533	%				
PP	552	90.8	494	92.7				
PL	54	8.9	36	6.8				
LL	2	0.3	3	0.6	0.35			
P (C)	1158	95.2	1024	96.1				
L (T)	58	4.8	42	3.9	0.33	1.22 (0.80-1.87)	17:19	0.44
R478W (rs5030400)	N=607	%	N=531	%				
RR	595	98.0	525	98.9				
RW	12	2.0	5	0.9				
WW	0	0.0	1	0.2	0.20			
R (C)	1202	99.0	1055	99.3				
W (T)	12	1.0	7	0.7	0.40	1.50 (0.55-4.23)	5:2	0.23

T = transmitted; U = untrasmitted

The familial study (TDT) did not demonstrate any differences either. No differences were observed after stratification by HLA-DQ2 status and by gender. No significant linkage disequilibrium was observed between any of the *ICAM1* polymorphisms and the *MYO9B* polymorphisms previously studied [5] (LOD<0.0010 for every pair of SNPs tested).

Table 2: Haplotypic frequencies of *ICAM1* SNPs (rs281432-rs1799969-rs1801714- rs5030400) in CD patients and controls, and TDT results.

	CD		Controls		P	OR (95% CI)	TDT	
							T:U	p
	2N=1215		2N=1080					
CGCC	523	43.0	494	45.6	0.20	0.95 (0.88-1.03)	106:123	0.15
GGCC	533	43.7	442	40.8	0.15	1.13 (0.95-1.34)	119:113	0.37
GACC	93	7.60	96	8.80	0.30	0.85 (0.63-1.16)	44:33	0.13
CGTC	55	4.50	41	3.80	0.41	1.20 (0.78-1.86)	17:20	0.40
CGCT	11	0.90	7	0.70	0.51	1.40 (0.50-4.01)	6:2	0.15

T = transmitted; U = untrasmitted

Discussion

We studied the role of polymorphisms in the *ICAM1* gene in CD susceptibility in the Spanish population. We analyzed the missense mutation G241R (rs1799969), previously studied in CD and in other autoimmune diseases, and additional polymorphisms with a potential functional role and demonstrating variability in our population. No association between CD and any of the SNPs studied existed. We could not replicate the previously described association between R241 and CD in the French population [11], despite our study being well enough powered. The statistical power of our study was above 99% when an odds ratio (OR) value of 2.9, as described by Abel *et al*. [11], is considered [and even as high as 98% when considering OR=1.7, which is the lower limit of the 95% confidence interval (CI)]. The described effect of this allele [11] seemed to be stronger in patients with an adulthood onset (OR = 4.2; 95% CI 2.3–7.5). Our CD patients were mainly pediatric (7% of our patients developed disease after 18 years old), which lowers the power to detect the R241 effect in this group of patients to 87%, which is still considered high power. Regarding the association described for pediatric patients (OR=2.1), we discard the effect with 99% power.

In conclusion, the putatively functional variation present in *ICAM1* does not seem to be involved in CD susceptibility in the Spanish population. Despite the numerous studies performed to find the gene/genes underlying the significant linkage to chromosome 19p13 [21], no consistently associated genes were reported. Besides

the study of *ICAM1* in the French population, to our knowledge an association in *CELIAC4* had been only convincingly described in the Dutch population, although a recent report described a strong linkage to but no association with *MYO9B* variants in a Finnish–Hungarian study [22]. Additional studies, including one meta-analysis using data from several European populations [23], did not demonstrate strong evidence of linkage to chromosome 19. Clearly, further studies are necessary to localize the relevant variant located in the 19p genomic region if it exists in all populations.

Acknowledgments

We are most grateful to Carmen Martínez Cuervo for her expert technical assistance. This work was supported by a grant from the Fundación Ramón Areces. Bárbara Dema received a grant from the Fundación Mutua Madrileña. Concepción Núñez and Alfonso Martínez have FIS contracts (CA06/0163 and CP04/00175, respectively) and Elena Urcelay works for the Fundación para la Investigación Biomédica-Hospital Clínico San Carlos. We declare no conflicts of interest.

References

[1] Monsuur AJ, de Bakker PI, Alizadeh BZ, Zhernakova A, Bevova MR, Strengman E, et al. Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. Nat Genet 2005;37:1341-4.

[2] Amundsen SS, Monsuur AJ, Wapenaar MC, Lie BA, Ek J, Gudjonsdottir H, et al. Association analysis of MYO9B gene polymorphisms with celiac disease in a Swedish/Norwegian cohort. Hum Immunol 2006;67:1341-5.

[3] Giordano M, Marano C, Mellai M, Limongelli MG, Bolognesi E, Clerget-Darpoux F, et al. A family-based study does not confirm the association of MYO9B with celiac disease in the Italian population. Genes Immun 2006;7:606-8.

[4] Hunt KA, Monsuur AJ, McArdle WL, Kumar PJ, Travis SP, Walters JR, et al. Lack of association of MYO9B genetic variants with coeliac disease in a British cohort. Gut 2006;55:969-72.

[5] Nunez C, Marquez A, Varade J, Martinez A, Polanco I, Maluenda C, et al. No evidence of association of the MYO9B polymorphisms with celiac disease in the Spanish population. Tissue Antigens 2006;68:489-92.

[6] Latiano A, Mora B, Bonamico M, Megiorni F, Mazzilli MC, Cucchiara S, et al. Analysis of candidate genes on chromosomes 5q and 19p in celiac disease. J Pediatr Gastroenterol Nutr 2007; 45:180-6.

[7] Curley CR, Monsuur AJ, Wapenaar MC, Rioux JD, Wijmenga C. A functional candidate screen for coeliac disease genes. Eur J Hum Genet 2006;14:1215-22.

[8] van de Stolpe A, van der Saag PT. Intercellular adhesión molecule-1. J Mol Med 1996;74:13-33.

[9] Sturgess RP, Macartney JC, Makgoba MW, Hung CH, Haskard DO, Ciclitira PJ. Differential upregulation of intercellular adhesion molecule-1 in coeliac disease. Clin Exp Immunol 1990;82:489-92.

[10] Jelinkova L, Tuckova L, Sanchez D, Krupickova S, Pozler O, Nevoral J, et al. Increased levels of circulating ICAM-1, Eselectin, and IL-2 receptors in celiac disease. Dig Dis Sci 2000; 45:398-402.

[11] Abel M, Cellier C, Kumar N, Cerf-Bensussan N, Schmitz J, Caillat-Zucman S. Adulthood-onset celiac disease is associated with intercellular adhesion molecule-1 (ICAM-1) gene polymorphism. Hum Immunol 2006;67:612-7.

[12] Kaur G, Rapthap CC, Kumar S, Bhatnagar S, Bhan MK, Mehra NK. Polymorphism in L-selectin, E-selectin and ICAM-1 genes in Asian Indian pediatric patients with celiac disease. Hum Immunol 2006;67:634-8.

[13] Nejentsev S, Laaksonen M, Tienari PJ, Fernandez O, Cordell H, Ruutiainen J, et al. Intercellular adhesion molecule-1 K469E polymorphism: study of association with multiple sclerosis. Hum Immunol 2003;64:345-9.

[14] Cournu-Rebeix I, Genin E, Lesca G, Azoulay-Cayla A, Tubridy N, Noe E, et al. Intercellular adhesion molecule-1: a protective haplotype against multiple sclerosis. Genes Immun 2003;4:518- 23.

[15] Nishimura M, Obayashi H, Maruya E, Ohta M, Tegoshi H, Fukui M, et al. Association between type 1 diabetes age-at-onset and intercellular adhesion molecule-1 (ICAM-1) gene polymorphism. Hum Immunol 2000;61:507-10.

[16] Nejentsev S, Guja C, McCormack R, Cooper J, Howson JM, Nutland S, et al. Association of intercellular adhesion molecule-1 gene with type 1 diabetes. Lancet 2003;362:1723-4.

[17] Matsuzawa J, Sugimura K, Matsuda Y, Takazoe M, Ishizuka K, Mochizuki T, et al. Association between K469E allele of intercellular adhesion molecule 1 gene and inflammatory bowel disease in a Japanese population. Gut 2003;52:75-8.

[18] Hong J, Leung E, Fraser AG, Merriman TR, Vishnu P, Krissansen GW. Polymorphisms in NFKBIA and ICAM-1 genes in New Zealand Caucasian Crohn's disease patients. J Gastroenterol Hepatol 2007;22:1666-70.

[19] Kretowski A, Wawrusiewicz N, Mironczuk K, Mysliwiec J, Kretowska M, Kinalska I. Intercellular adhesion molecule 1 gene polymorphisms in Graves' disease. J Clin Endocrinol Metab 2003;88:4945-9.

[20] Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. Arch Dis Child 1990;65:909-11.

[21] Van Belzen MJ, Meijer JW, Sandkuijl LA, Bardoel AF, Mulder CJ, Pearson PL, et al. A major non-HLA locus in celiac disease maps to chromosome 19. Gastroenterology 2003;125:1032-41.

[22] Koskinen LL, Korponay-Szabo IR, Viiri K, Juuti-Uusitalo K, Kaukinen K, Lindfors K, et al. Myosin IXB gene region and gluten intolerance: linkage to coeliac disease and a putative dermatitis herpetiformis association. J Med Genet 2008;45:222-7.

[23] Babron MC, Nilsson S, Adamovic S, Naluai AT, Wahlstrom J, Ascher H, et al. Meta and pooled analysis of European coeliac disease data. Eur J Hum Genet 2003;11:828-34.

II. Lack of replication of celiac disease risk variants reported in a Spanish population using an independent Spanish sample

Bárbara Dema[a], Alfonso Martínez[a], Miguel Fernández-Arquero[a], Carlos Maluenda[b], Isabel Polanco[c], Emilio G. de la Concha[a], Elena Urcelay[a], Concepción Núñez[a]

[a] Servicio de Inmunología Clínica, Hospital Clínico San Carlos, 28040 Madrid, Spain
[b] Servicio de Pediatría, Hospital Clínico San Carlos, 28040 Madrid, Spain
[c] Servicio de Gastroenterología Pediátrica, Hospital La Paz, 28046 Madrid, Spain

Genes and Immunity (2009); 10(7):659-61

Resumen

La enfermedad celíaca (EC) es una condición inflamatoria que afecta al intestino delgado y se desencadena en individuos genéticamente susceptibles al ingerir gluten (o proteínas relacionadas). Recientemente, se han visto asociados a EC en población española, polimorfismos en tres genes, *SERPINE2*, *PPP6C* y *PBX3*. Sin embargo, esta asociación no se ha replicado en población inglesa utilizando datos imputados. Como este segundo estudio analizaba una población distinta, nos planteamos reevaluar el papel de estos polimorfismos con una muestra española independiente. Genotipamos tres polimorfismos de un único nucleótido: rs6747096 en *SERPINE2*, rs458046 en *PPP6C* y rs7040561 en *PBX3*, en 417 pacientes celíacos, 527 controles sanos étnicamente pareados y en ambos progenitores de 304 enfermos celíacos. Se llevó a cabo un estudio caso-control utilizando el test χ^2 y un estudio familiar con el test de desequilibrio de transmisión. No detectamos asociación en dichos análisis. Por tanto, nuestros resultados parecen descartar el papel previamente descrito de los polimorfismos en *SERPINE2*, *PPP6C* y *PBX3* en la susceptibilidad a EC.

Abstract

Celiac disease (CD) is an inflammatory condition affecting small bowel and triggered by gluten (or related proteins) ingestion in genetic susceptible individuals. Polymorphisms in three genes, *SERPINE2*, *PPP6C* and *PBX3*, have recently been associated with CD in the Spanish population. However, this association could not be replicated in the UK population using imputed data. As this second study analyzed a different population, we aimed at reevaluating the role of those polymorphisms using an independent Spanish sample. We genotyped three single nucleotide polymorphisms: rs6747096 in *SERPINE2*, rs458046 in *PPP6C* and rs7040561 in *PBX3*, in 417 CD patients, 527 ethnically matched healthy controls and parents of 304 CD patients. A case–control study using the χ^2-test and a familial study using the transmission disequilibrium test were performed. No association was detected in those analyses. Therefore, our results seem to discard the role of the previously described polymorphisms in *SERPINE2*, *PPP6C* and *PBX3* in CD susceptibility.

KEYWORDS: *SERPINE2*; *PPP6C*; *PBX3*; CD susceptibility.

Introduction

Celiac disease (CD) is a complex disease mediated by immune processes triggered after ingestion of gluten or related proteins in genetically susceptible individuals [1]. The genetic basis of this disease is being slowly unraveled by advances in experimental techniques. Combining two different gene-search approaches, Castellanos-Rubio et al. [2] reported new CD risk variants in the Spanish Basque population. They combined information provided by whole-genome expression profiling experiments and linkage studies (that is, functional and positional information) and found significant association with four single nucleotide polymorphisms (SNPs) located in the *SERPINE2* (serine protease inhibitor, clade E, member 2), *PPP6C* (protein phosphatase 6, catalytic subunit) and *PBX3* (pre-B-cell leukemia homeobox 3) genes.

The *SERPINE2* gene maps in the CD linkage region 2q33–q35 and *PPP6C* and *PBX3* in 9q33–q34. *SERPINE2* is involved in extracellular matrix production and it has been widely studied in relation to COPD (chronic obstructive pulmonary disease) [3]. *PPP6C* encodes a protein phosphatase which seems to be involved in cell-cycle regulation [4]. Finally, PBX3 is a transcription factor implicated in basic developmental functions, including some related to immune cells [5].

The association between those genes and CD susceptibility described by Castellanos-Rubio et al. was questioned by Hunt et al. [6], who used data from a genomewide association study performed in the British population to impute data corresponding to the previously associated SNPs. No replication was obtained and the authors suggested different possible explanations. Therefore, the debate is open because for every SNP the imputation is not 100% accurate, an issue recently discussed [7]. As an explanation of the observed discrepancies between the two studies, could be owed to population or clinical heterogeneity, we aimed at evaluating the role of those SNPs in an independent Spanish sample of pediatric typical CD patients.

Materials and Methods

We studied 417 unrelated Spanish CD patients (diagnosed following the European Society for Paediatric Gastroenterology and Nutrition (ESPGAN) [173] criteria), 527 ethnically matched healthy controls (mainly blood donors and hospital staff) and parents of 304 out of those CD patients. A total of 97% CD patients are HLA-DQ2 and/or HLA-DQ8 positive. Samples were consecutively collected in two centres of the same region (Hospital La Paz and Hospital Clínico San Carlos, Madrid). A written informed consent was obtained from all the participants and the Ethics Committee of the Hospital Clínico San Carlos approved this study.

All samples were genotyped by TaqMan technology following manufacturer recommendations (Applied Biosystems Inc., Foster City, CA, USA) for the three SNPs previously reported as involved in disease susceptibility: rs6747096 in *SERPINE2*, rs458046 in *PPP6C* and rs7040561 in *PBX3*. We did not include the also significantly associated SNP rs459311 in *PPP6C*, due to its high correlation with the already included rs458046 (r^2=0.97).

Case-control comparisons were performed using the chi-square test; and a transmission disequilibrium test (TDT) was used in the familial analyses. Statistical power calculations were carried out with the EpiInfo v6.02 software.

Results and Discussion

Allele frequencies obtained in our case–control study are shown in Table 1 together with the previously published data [2,6]. No significant differences emerged when comparing our CD patients and controls. Additionally, we performed a transmission disequilibrium test (TDT) with data from 304 trios and no significant association was observed either.

As the two SNPs analyzed in the 9q region are in moderate linkage disequilibrium (D'=0.5, r^2=0.037, in our control sample), haplotypic analysis was performed, but significance was not improved (data not shown).

Therefore, our results seem to confirm the lack of association with CD of the studied SNPs in the *SERPINE2*, *PPP6C* and *PBX3* genes, concordantly with the data reported by Hunt et al.[6]. This negative result is probably not due to lack of statistical power, because our study shows more than 99.9% statistical power (calculated with EpiInfo v.6.02, World Health Organization, Geneva, Switzerland) to detect the effects (odds ratios, ORs) initially described (at p=0.01). The risks originally reported could be overestimated, but we can reach 80% statistical power to detect modest ORs (0.7, 1.3 and 1.5 in the *SERPINE2*, *PPP6C* and *PBX3* genes, respectively). Nonetheless, smaller-size effects cannot be formally discarded. Differences in the clinical features of the patients studied probably do not affect; we used a uniform Spanish population composed of only pediatric (mean age=6.0 years (range, 9 months to 17 years)) and typical CD patients, to minimize differences with the previously studied Spanish group. Interestingly, our TDT study confirms the lack of association precluding population stratification, a problem that can be present in case–control studies and, therefore, might affect the Castellanos-Rubio et al. Study [2]. When some stratification is present in the studied population, a different proportion of individuals belonging to diverse ethnic subgroups in cases and controls may lead to significant differences in allelic frequencies between cases and controls, but this problem is avoided with the TDT. Therefore, a TDT study in the Basque population would be very interesting, as it could definitively solve this issue. Provided the positive association is confirmed, a possible specific effect in the Basque population could be discovered.

Although the *SERPINE2* polymorphism is not associated with CD, this does not mean that the widely replicated CD linkage region 2q33 (*CELIAC3*) is not a true susceptibility locus. This region also contains the candidate *CTLA4/ICOS* genes, approximately 20Mb apart, extensively studied in relation to CD due to their role in T-lymphocyte activity regulation, and recently associated with CD susceptibility in a population with proven linkage [8]. The *PPP6C* and *PBX3* loci are located in 9q. This region showed linkage to CD in the North American population [9] but, in contrast to 2q33, no replication has been reported. Further studies are necessary to confirm this possible linkage region and, if replicated, to look for the etiologic genes.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We are most grateful to Carmen Martínez Cuervo and M.Ángel García Martínez for their expert technical assistance. This work was supported by project CP08/0213 from 'Fondo de Investigaciones Sanitarias'. B. Dema received a grant from 'Fundación Mutua Madrileña'. C. Núñez and A. Martínez have a FIS contract (CP08/0213 and CP04/00175, respectively) and E. Urcelay works for the 'Fundación para la Investigación Biomédica-Hospital Clínico San Carlos'.

Table 1 Allelic frequencies of the three studied SNPs

Gene	SNP	Alleles	Spanish (Castellanos-Rubio et al)			UK GWAS			Spanish replication sample				
									Case-control			TDT	
		1/2	1 in cases	1 in controls	p[a]	1 in cases	1 in controls	p	1 in cases	1 in controls	p	1T:1U	p
SERPINE2	rs6747096	G/A	0.16	0.29	0.008	0.21	0.18	0.016[b]	0.24	0.22	0.56	106:100	0.36
PPP6C	rs458046	T/A	0.52	0.39	0.043	0.43	0.43	0.21	0.47	0.46	0.84	148:141	0.36
PBX3	rs7040561	T/A	0.15	0.06	0.021	0.16	0.14	0.12	0.10	0.11	0.57	50:69	0.05

a P values correspond to the corrected values after using Bonferroni correction.
b Note that this borderline significant result is in the opposite direction than the originally described.

References

[1] van Heel DA, West J. Recent advances in coeliac disease. Gut 2006; 55: 1037–1046.

[2] Castellanos-Rubio A, Martin-Pagola A, Santin I, Hualde I, Aransay AM, Castano L et al. Combined functional and positional gene information for the identification of susceptibility variants in celiac disease. Gastroenterology 2008; 134:738–746.

[3] Seifart C, Plagens A. Genetics of chronic obstructive pulmonary disease. Int J Chron Obstruct Pulmon Dis 2007; 2: 541–550.

[4] Bastians H, Ponstingl H. The novel human protein serine/ threonine phosphatase 6 is a functional homologue of budding yeast Sit4p and fission yeast ppe1, which are involved in cell cycle regulation. J Cell Sci 1996; 109(Pt 12):2865–2874.

[5] Penkov D, Di Rosa P, Fernandez Diaz L, Basso V, Ferretti E, Grassi F et al. Involvement of Prep1 in the alphabeta T-cell receptor T-lymphocytic potential of hematopoietic precursors. Mol Cell Biol 2005; 25: 10768–10781.

[6] Hunt KA, Franke L, Deloukas P, Wijmenga C, van Heel DA. No evidence in a large UK collection for celiac disease riskvariants reported by a Spanish study. Gastroenterology 2008; 134: 1629–1630. author reply 1630-1621.

[7] Rao DC. An overview of the genetic dissection of complex traits. Adv Genet 2008; 60: 3–34.

[8] Haimila K, Einarsdottir E, de Kauwe A, Koskinen LL, Pan-Hammarstrom Q, Kaartinen T et al. The shared CTLA4-ICOS risk locus in celiac disease, IgA deficiency and common variable immunodeficiency. Genes Immun 2008; 10:151–161.

[9] Garner CP, Ding YC, Steele L, Book L, Leiferman K, Zone JJ et al. Genome-wide linkage analysis of 160 North American families with celiac disease. Genes Immun 2007; 8:108–114

[10] Revised criteria for diagnosis of coeliac disease. Report of working group of european society of paediatric gastroenterology and nutrition. Arch Dis Child 1990; 65: 909–911.

III. Association of *IL18RAP* and *CCR3* with celiac disease in the Spanish population

Bárbara Dema[a], Alfonso Martínez[a], Miguel Fernández-Arquero[a], Carlos Maluenda[b], Isabel Polanco[c], Emilio G. de la Concha[a], Elena Urcelay[a], Concepción Núñez[a]

[a] Servicio de Inmunología Clínica, Hospital Clínico San Carlos, 28040 Madrid, Spain
[b] Servicio de Pediatría, Hospital Clínico San Carlos, 28040 Madrid, Spain
[c] Servicio de Gastroenterología Pediátrica, Hospital La Paz, 28046 Madrid, Spain

Journal of Medical Genetics (2009); 46 (9): 617-9

Resumen

Antecedentes y objetivos: los estudios de asociación por barrido genómico en enfermedad celíaca (EC) mostraron ocho nuevas regiones asociadas a la susceptibilidad a la enfermedad. Sin embargo, un estudio de replicación llevado a cabo en población italiana no pudo confirmar dos de estas regiones: 2q12 (*IL18RAP*) y 3p21 (*CCR3*). Nos planteamos el estudio del papel de estas regiones en el riesgo a padecer EC, en una población mediterránea diferente, la población española.

Métodos: se llevó a cabo un estudio caso-control con 722 pacientes con EC y 794 controles sanos étnicamente pareados. Dos polimorfismos de un único nucleótido, rs917997 (2q12) y rs6441961 (3p21), fueron genotipados y las frecuencias génicas entre ambos grupos fueron comparadas con el test chi-cuadrado.

Resultados: se encontró asociación con el polimorfismo rs6441961: p=0.0004, OR=1.32 95% CI 1.13-1.54. Cuando se estudió el polimorfismo rs917997, se obtuvo un resultado no significativo pero concordante al del estudio inicial.

Conclusión: confirmamos la asociación de la región 3p21 con la susceptibilidad a EC en población española. En 2q12, la OR inicialmente descrita se encuentra probablemente sobreestimada, y por tanto la situación real puede ser la existencia de un factor genuino pero débil, que genera limitaciones de potencia estadística.

Abstract

Background and aims: genome wide association studies in celiac disease (CD) showed eight new genetic regions associated with disease susceptibility. However, a replication study performed in the Italian population could not confirm two of those new regions: 2q12 (*IL18RAP*) and 3p21 (*CCR3*). We aimed to study the role of those regions in CD risk in a different Mediterranean population, the Spanish one.

Methods: a case-control study with 722 CD patients and 794 ethnically matched healthy controls was performed. Two single nucleotide polymorphisms, rs917997 (2q12) and rs6441961 (3p21), were genotyped and their genetic frequencies were compared between both groups with the chi-square test.

Results: association was found with rs6441961: p=0.0004, OR=1.32 95% CI 1.13-1.54. A non-significant (but concordant with the initial study) result was obtained when studying rs917997.

Conclusion: we confirmed the association of the 3p21 genetic region with CD susceptibility in the Spanish population. In 2q12, the initially described OR is most probably overestimated, and therefore the real situation may be the existence of a genuine but weak risk factor, which generates statistical power limitations.

KEYWORDS: *IL18RAP*; *CCR3*; replication, celiac disease susceptibility.

Introduction

Celiac disease (CD) is a chronic inflammatory disorder affecting the small bowel and triggered by ingestion of specific proteins present mainly in wheat, rye and barley, in genetic susceptible individuals. It is a common disease, with a prevalence of approximately 1 per 100 in individuals of European ancestry [5]. Several alleles of the *HLA* region (those encoding the HLA-DQ2/DQ8 heterodimers) were, for a long time, the only firmly established genetic factors contributing to CD risk. However, last years have witnessed the finding of new CD susceptibility genes propelled by the development of genome wide association studies (GWAS). Eight new genetic regions have now been associated and replicated with CD susceptibility in several Northern European populations [126, 132]. All those regions contain potential candidate gene/s, most of them related with the immune response. In contrast, a recent study using Italian individuals [135] failed to replicate the association with two of those genetic regions, 2q12 and 3p21, harbouring the genes *IL18RAP* and *CCR3*, respectively. The authors suggest different possibilities for discrepancy as clinical heterogeneity, lack of statistical power or use of inappropriate markers for capturing the causal variant, but also claim that a genuine population difference of CD susceptibility loci across Europe could exist. Therefore, we wanted to analyse the role of those SNPs which did not show association with CD risk in the Italian population in a different Southern European population, as the Spanish one. With that aim, we genotyped the SNPs located in the 2q12 and 3p21 regions showing the most significant association with CD in the initial studies, in Spanish CD patients and controls.

Material and Methods

We studied 722 CD patients and 794 ethnically matched healthy controls. CD patients were diagnosed following the European Society for Paediatric Gastroenterology and Nutrition (ESPGAN), 62% are female and 97% are positive for HLA-DQ2 and/or HLA-DQ8. Most of our CD patients were pediatric (mean age = 7.4 years [range, 9 months to 67 years]) and suffer from typical CD. Controls correspond mainly to blood donors and laboratory staff. CD samples were consecutively collected in two centres of the same region (Hospital La Paz and Hospital Clínico San Carlos, Madrid) and controls were collected at the Hospital Clínico San Carlos. A written

informed consent was obtained from all the participants and the Ethics Committee of the Hospital Clínico San Carlos approved this study.

DNA was extracted from fresh peripheral blood leukocytes by a "salting out" procedure. All samples were genotyped by TaqMan technology following manufacturer recommendations (Applied Biosystems Inc., Foster City, USA) for two single nucleotide polymorphisms (SNPs): rs917997 in the 2q12 region, downstream of the *IL18RAP* gene and rs6441961 in 3p21. These two SNPs were selected because they showed the strongest association with CD in the GWAS previously performed [126, 132] and because they were selected as the SNPs to be replicated in the Italian study.[135] Success genotyping rate was over 99% for both SNPs.

Case-control comparisons were performed with the chi-square test using the statistical package EpiInfo v5.00 (CDC, Atlanta, USA), which was also used for statistical power calculations. A meta-analysis using published data [134] was performed with Review Manager (RevMan) 5.0 software (Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2008), which uses the Mantel-Haenszel test for calculating combined ORs.

Results

The genotype distributions of the SNPs studied in this work conformed to Hardy-Weinberg expectations.

Genotypic and allelic frequencies are shown in Table 1. No significant differences between patients and controls are observed when studying the SNP located in the 2q11-q12 region. However, a highly significant difference between both groups emerged for rs6441961.

Table 1: Genotypic and allelic frequencies of the two SNPs analyzed in CD patients and controls.

Chromosomal region	SNP	CD patients	Controls	p	OR (95% CI)
		n (%)	n (%)		
2q11-q12	**rs917997**	**n=722**	**n=794**		
	GG*	370 (51.2)	426 (53.7)	-	1.00
	GA	293 (40.6)	315 (39.7)	0.53	1.07 (0.86-1.33)
	AA	59 (8.17)	53 (6.67)	0.22	1.28 (0.85-1.94)
	G	1033 (71.5)	1167 (73.5)		
	A	411 (28.5)	421 (26.5)	0.23	1.10 (0.94-1.30)
3p21	**rs6441961**	n=722	n=790		
	GG*	298 (41.3)	398 (50.4)	-	1.00
	GA	344 (47.6)	327 (41.4)	0.002	1.41 (1.13-1.75)
	AA	80 (11.1)	65 (8.23)	0.007	1.64 (1.13-2.39)
	G	940 (65.1)	1123 (71.1)		
	A	504 (34.9)	457 (28.9)	0.0004	1.32 (1.13-1.54)

*This genotype was taken as the reference for genotypic calculations.

Stratifications by HLA-DQ2 status or by sex were performed but no significant results were observed.

The results of the meta-analysis using the previously negative published data and our present data are shown in Figure 1. A borderline significance is obtained.

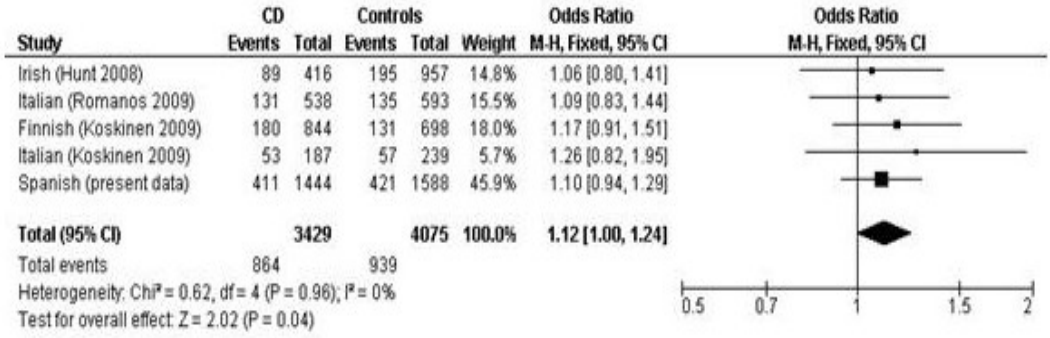

Figure 1: Results of the meta-analysis for *IL18RAP* rs917997 and CD using previous negative studies and our present data.

Discussion

GWAS in CD resulted in eight new genetic regions associated with disease susceptibility. However, two of these regions, 2q12 and 3p21, could not be subsequently replicated in the Italian population. In this study, we have investigated the involvement of SNPs in those two regions in CD risk in the Spanish population.

A significant association with CD was found when studying the role of the polymorphism rs6441961, located in the 3p21 genetic region. That region contains a cluster of chemokine receptor genes, being *CCR3* (chemokine (C-C motif) receptor 3), located 44 kb apart of the studied SNP, the nearest gene. This finding is functionally relevant since chemokines are involved in leukocyte trafficking during infection or inflammation and, therefore, chemokines and their receptors have been previously implicated in several pathogenic conditions [141-143]. In CD, chemokine receptors, specifically CXCR3, has been even related to gliadin binding and increased intestinal permeability.[26] Further studies must be now directed toward the specific functional implication of this finding.

When studying the SNP located downstream of the *IL18RAP* gene, we did not detect a significant association. Although this result is concordant with the one observed in the Italian population [132], the lack of association could be due to a problem of statistical power. A study of *IL18RAP* in three populations: Finnish, Hungarian and Italian [134], has been recently published. A significant association was only found with the Hungarian samples, but all populations showed a similar trend, and a meta-analysis confirmed the involvement of the *IL18RAP* studied polymorphisms (rs13015714 and rs917997) in CD susceptibility. The dose effect detected in that study can be observed in our Spanish sample (although non-significant: rs917997_GA showed an OR=1.07; and rs917997_AA showed an OR=1.28). Moreover, when a meta-analysis is performed adding our Spanish data to the negative previously published data (from two Italian, one Finnish and one Irish populations), borderline significance is obtained: OR=1.12 95% CI 1.00-1.24, p=0.044. Most probably, the effect described in the original study, OR=1.34 95% CI 1.16-1.54 [132], was slightly overestimated. Consequently, statistical power limitations exist and a lack of association is observed. The new data suggest an OR close to 1.22, which results in

only a moderate statistical power (68%) in the present study. Although, in this situation, the clinical relevance of this finding could be arguable, it should be noted that that low OR could be the consequence of an important role of *IL18RAP* polymorphisms in a small subset of patients which could be characterized by the simultaneous presence of several risk factors (showing low ORs when considered individually). Alternatively, a low OR could also be a mere spurious consequence of stratification in the populations analysed; however, we performed a TDT study (which is immune to population substructure) with 309 families and a slight (although statistically non-significant) over-transmission of the allele A was also observed: 131 transmitted vs. 112 non-transmitted. *IL18RAP* (interleukin 18 receptor accessory protein) is necessary for IL-18 signalling and, therefore, could be important in Th1-mediated diseases. In addition, functional studies have been performed and rs917997 genotypes seem to correlate with the level of *IL18RAP* mRNA levels in whole blood from treated CD individuals [132]. Again, further investigations are required; it must be remembered that *IL18RAP*, as well as *CCR3*, were found associated with CD in GWAS, these studies are hypothesis-free and, therefore, the functional implication of the observed SNPs and the disease need more in depth analyses.

In conclusion, our work seems to confirm the involvement of *IL18RAP* and, more clearly, of *CCR3* in CD risk also in a Southern European population. Moreover, this finding is also noticeable since they can represent susceptibility factors shared by several immune-mediated diseases: *CCR3* has been associated with Kawasaki disease [182] and pollinosis [102] and *IL18RAP* with Crohn´s disease [139] and asthma.[183]

Key points

1. Genome wide association studies showed eight new genetic regions associated with celiac disease, however two of them, 2q12 (*IL18RAP*) and 3p21 (*CCR3*), could not be replicated in an Italian population.

2. We confirmed the association of the single nucleotide polymorphism rs6441961, located near the *CCR3* gene, with celiac disease susceptibility in the Spanish population.

3. A non-significant association was observed with rs917997 (*IL18RAP*) but probably due to statistical power limitations, as the meta-analysis of the present data with previously negative published results evidenced a significant overall effect (OR=1.12).

Acknowledgements

We are most grateful to Carmen Martínez Cuervo and M. Ángel García Martínez for their expert technical assistance.

Competing interests

No conflicts of interests.

Funding

This work was supported by project CP08/0213 from "Fondo de Investigaciones Sanitarias". Bárbara Dema receives a grant from "Fundación Mutua Madrileña". Concepción Núñez and Alfonso Martínez have a FIS contract (CP08/0213 and CP04/00175, respectively) and Elena Urcelay works for the "Fundación para la Investigación Biomédica-Hospital Clínico San Carlos".

Licence for publication

References

[1] van Heel DA, West J. Recent advances in coeliac disease. *Gut* 2006;**55**(7):1037-46.

[2] van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, Wapenaar MC, Barnardo MC, Bethel G, Holmes GK, Feighery C, Jewell D, Kelleher D, Kumar P, Travis S, Walters JR, Sanders DS, Howdle P, Swift J, Playford RJ, McLaren WM, Mearin ML, Mulder CJ, McManus R, McGinnis R, Cardon LR, Deloukas P, Wijmenga C. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* 2007;**39**(7):827-9.

[3] Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M, Romanos J, Dinesen LC, Ryan AW, Panesar D, Gwilliam R, Takeuchi F, McLaren WM, Holmes GK, Howdle PD, Walters JR, Sanders DS, Playford RJ, Trynka G, Mulder CJ, Mearin ML, Verbeek WH, Trimble V, Stevens FM, O'Morain C, Kennedy NP, Kelleher D, Pennington DJ, Strachan DP, McArdle WL, Mein CA, Wapenaar MC, Deloukas P, McGinnis R, McManus R, Wijmenga C, van Heel DA. Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet* 2008;**40**(4):395-402.

[4] Romanos J, Barisani D, Trynka G, Zhernakova A, Bardella MT, Wijmenga C. Six new coeliac disease loci replicated in an Italian population confirm association with coeliac disease. *J Med Genet* 2009;**46**(1):60-3.

[5] Koskinen LL, Einarsdottir E, Dukes E, Heap GA, Dubois P, Korponay-Szabo IR, Kaukinen K, Kurppa K, Ziberna F, Vatta S, Not T, Ventura A, Sistonen P, Adany R,

Pocsai Z, Szeles G, Maki M, Kere J, Wijmenga C, van Heel DA, Saavalainen P. Association study of the IL18RAP locus in three European populations with coeliac disease. *Hum Mol Genet* 2009;**18**(6):1148-55.

[6] Papadakis KA. Chemokines in inflammatory bowel disease. *Curr Allergy Asthma Rep* 2004;**4**(1):83-9.

[7] Mamtani M, Rovin B, Brey R, Camargo JF, Kulkarni H, Herrera M, Correa P, Holliday S, Anaya JM, Ahuja SK. CCL3L1 gene-containing segmental duplications and polymorphisms in CCR5 affect risk of systemic lupus erythaematosus. *Ann Rheum Dis* 2008;**67**(8):1076-83.

[8] McKinney C, Merriman ME, Chapman PT, Gow PJ, Harrison AA, Highton J, Jones PB, McLean L, O'Donnell JL, Pokorny V, Spellerberg M, Stamp LK, Willis J, Steer S, Merriman TR. Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. *Ann Rheum Dis* 2008;**67**(3):409-13.

[9] Lammers KM, Lu R, Brownley J, Lu B, Gerard C, Thomas K, Rallabhandi P, Shea-Donohue T, Tamiz A, Alkan S, Netzel-Arnett S, Antalis T, Vogel SN, Fasano A. Gliadin induces an increase in intestinal permeability and zonulin release by binding to the chemokine receptor CXCR3. *Gastroenterology* 2008;**135**(1):194-204 e3.

[10] Breunis WB, Biezeveld MH, Geissler J, Kuipers IM, Lam J, Ottenkamp J, Hutchinson A, Welch R, Chanock SJ, Kuijpers TW. Polymorphisms in chemokine receptor genes and susceptibility to Kawasaki disease. *Clin Exp Immunol* 2007;**150**(1):83-90.

[11] Nakamura H, Higashikawa F, Nobukuni Y, Miyagawa K, Endo T, Imai T, Hatta K, Ozasa K, Motohashi Y, Matsuzaki I, Sasahara S, Ogino K, Akimaru K, Eboshida A. Genotypes and haplotypes of CCR2 and CCR3 genes in Japanese cedar pollinosis. *Int Arch Allergy Immunol* 2007;**142**(4):329-34.

[12] Zhernakova A, Festen EM, Franke L, Trynka G, van Diemen CC, Monsuur AJ, Bevova M, Nijmeijer RM, van 't Slot R, Heijmans R, Boezen HM, van Heel DA, van Bodegraven AA, Stokkers PC, Wijmenga C, Crusius JB, Weersma RK. Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring CARD9 and IL18RAP. *Am J Hum Genet* 2008;**82**(5):1202-10.

[13] Reijmerink NE, Postma DS, Bruinenberg M, Nolte IM, Meyers DA, Bleecker ER, Koppelman GH. Association of IL1RL1, IL18R1, and IL18RAP gene cluster polymorphisms with asthma and atopy. *J Allergy Clin Immunol* 2008;**122**(3):651-4 e8.

2.- ESTUDIO DE GENES ASOCIADOS A OTRAS ENFERMEDADES AUTOINMUNES

I. *IL23R*: a susceptibility locus for celiac disease and multiple sclerosis?

II. Lack of association of *NKX2-3*, *IRGM*, and *ATG16L1* inflammatory bowel disease susceptibility variants with celiac disease.

III. Autoimmune disease association signals in *CIITA* and *KIAA0350* are not involved in celiac disease susceptibility.

I. *IL23R*: a susceptibility locus for celiac disease and multiple sclerosis?

C Núñez [a], B Dema [a], MC Cénit [a], I Polanco [b], C Maluenda [c], R Arroyo [d], V de las Heras [d], M Bartolomé [d], EG de la Concha [a], E Urcelay [a] and A Martínez [a]

[a] Servicio de Inmunología Clínica, Hospital Clínico San Carlos, 28040 Madrid, Spain
[b] Servicio de Gastroenterología Pediátrica, Hospital La Paz, 28046 Madrid, Spain
[c] Servicio de Pediatría, Hospital Clínico San Carlos, 28040 Madrid, Spain
[d] Unidad de Esclerosis Múltiple, Hospital Clínico San Carlos, 28040 Madrid, Spain

Genes and Immunity (2008); 9 (4): 289-293

Resumen

Estudios recientes han mostrado asociación del gen *IL23R* con enfermedad inflamatoria intestinal, psoriasis y espondilitis anquilosante. Nos planteamos el estudio de la implicación de *IL23R* en enfermedad celíaca (EC) y esclerosis múltiple (EM). Llevamos a cabo un estudio caso-control incluyendo 598 pacientes con EC, 414 con EM y 546 controles sanos, todos ellos españoles. Todas las muestras fueron genotipadas para dos polimorfismos de un único nucleótido: rs7517847 y rs11209026 (Arg381Gln). Los análisis estadísticos se llevaron a cabo utilizando el test χ^2 o el test exacto de Fisher. El alelo minoritario (Gln) de la variante codificante Arg381Gln se encontraba significativamente aumentado en pacientes con EC y EM, al compararlos con los controles (8% en EC vs 6% en controles, p=0.02; 9% en EM, p=0.006). En EM, se observó un mayor efecto en pacientes con la forma primaria-progresiva de la enfermedad (16%, p=0.004). Además, los heterocigotos para el polimorfismo rs7517847 estaban significativamente aumentados en este grupo de pacientes con EM (81% en EM vs 48% en controles, p=0.0002). En conclusión, al contrario que lo descrito previamente, el alelo menos frecuente del polimorfismo funcional Arg381Gln (rs11209026), parece aumentar la susceptibilidad a EC y EM, aunque en este último grupo de pacientes se observa un efecto más fuerte en los pacientes afectados con la forma primaria-progresiva.

Abstract

Recent studies have shown association of the *IL23R* gene with inflammatory bowel disease, psoriasis and ankylosing spondylitis. We aimed at studying the involvement of *IL23R* in celiac disease (CD) and multiple sclerosis (MS). We performed a case–control study including 598 patients with CD, 414 with MS and 546 healthy controls, all of them white Spaniards. All samples were genotyped for two single nucleotide polymorphisms: rs7517847 and rs11209026 (Arg381Gln). Statistical analyses were performed using χ^2-tests or the Fisher's exact test. The minor allele (Gln) of the coding variant Arg381Gln was significantly increased in CD and MS patients when compared to controls (8% in CD vs 6% in controls, p=0.02; 9% in MS, p=0.006). In MS, a stronger effect was observed in patients showing primary-progressive disease (16%, p=0.004). Moreover, heterozygotes for rs7517847 were significantly increased in this group of MS patients (81% in MS vs 48% in controls, p=0.0002). In conclusion, contrary to what has been described previously, the less frequent allele of the functional polymorphism Arg381Gln (rs11209026) seems to be increasing susceptibility to CD and MS, although in this last group of patients a stronger effect is observed in patients affected of a primary-progressive form.

KEYWORDS: interleukin-23; IL23R; celiac disease; multiple sclerosis; single nucleotide polymorphism.

Introduction

The discovery in 2000 [1] of the new cytokine subunit p19, which together with p40 constitutes interleukin (IL)-23, has forced to reanalyze in last years many of the roles previously attributed to IL-12. Both cytokines, IL-23 and IL-12, are comprised by two subunits: one specific, p19 for IL-23 and p35 for IL-12, and one shared, p40. Since most roles of IL-12 had been addressed based on data regarding the p40 subunit, IL-23 could also be responsible of several functions traditionally attributed to IL-12. In fact, new studies capable of distinguishing between the roles of both cytokines seemed to indicate that IL-23 rather than IL-12 was the key relevant molecule in inflammation. Both cytokines also bind to heterodimeric receptors with unique, IL-23R for IL-23 and IL-12Rβ2 for IL-12, and shared, IL-12Rβ1, subunits. The discovery of IL-23 has added new complexity to the developmental options open to T cells: the two traditionally considered subsets of T helper (Th) cells, Th1 and Th2, mainly characterized by production of interferon (IFN)-γ and IL-4, respectively, must fit with the existence of T cells producing mainly IL-17 in response to IL-23 or IL-1β (human Th17 cells) [2].

Recent large-scale genetic scans have shown association of several polymorphisms located in the gene coding IL-23R (*IL23R*) with different inflammatory conditions as Crohn's disease, psoriasis or ankylosing spondylitis [3–5]. These conditions had been traditionally considered as Th1 diseases and consequently, an important role of IL-12 was expected. However, these new findings seem to indicate a more relevant role of IL-23 on these pathologies. The *IL23R* coding variant Arg381Gln (rs11209026) was associated with disease, with its minor allele showing protection (381Gln). Although additional variants have been described as associated, probably most of them are not independent (as many of the pairwise r^2-values between the single nucleotide polymorphisms (SNPs) found to be associated with Crohn's disease [3] are above 0.3).

Multiple sclerosis (MS) and celiac disease (CD) are two inflammatory conditions traditionally considered as typical Th1 diseases, mainly due to the high levels of IFN-γ mRNA observed [6,7]. We aimed at analyzing the involvement of *IL23R* polymorphisms in these two inflammatory conditions. Additionally, we

included in these analyses several polymorphisms of the gene coding for the IL-12p40 subunit (*IL12B*). Since this pair of molecules is in the same pathway, the disease phenotype can be either the consequence of mutations present simultaneously in both genes or, alternatively, of mutations present in one or the other gene, and in both cases a statistical interaction could be observed.

Materials and Methods

Subjects

We studied 598 patients with CD, 414 patients with MS and 546 ethnically matched controls. All samples corresponded to Spanish white individuals and were recruited from the Madrid area (Spain). CD was diagnosed following the European Society for Paediatric Gastroenterology and Nutrition criteria [8] and MS following the Poser criteria [9]. Controls were mainly anonymous blood donors and laboratory staff. A total of 92% of CD patients were HLA-DQ2 positive (*DQA1*05* and *DQB1*02*) and 36% of MS patients were *HLA-DRB1*1501* positive. In total 80% of MS patients were affected by the primary progressive (PP) form.Written informed consent was obtained from all subjects. In our control group, 30% of the individuals are HLA-DQ2 positive and 17% positive for *HLA-DRB1*1501*. This study was approved by the ethics committee of the Hospital Clínico San Carlos (Madrid).

Genotyping

We studied two SNPs on the *IL23R* gene, on chromosome 1: the intronic polymorphism rs7517847 and the coding variant rs11209026, which codes the amino-acid change Arg381Gln. Moreover, two additional SNPs around the *IL12B* gene were analyzed: rs6887695 and rs3212227. All SNPs were genotyped by TaqMan technology under conditions recommended by the manufacturer (Applied Biosystems, Foster City, CA, USA).

Statistical analyses

Genotypic, allelic and haplotypic frequency comparisons were performed using the χ^2-test or Fisher's exact test when expected values were below 5. Haplotypic frequencies were estimated using the Expectation-Maximization algorithm implemented in the Arlequin v2.000 software, with number of iterations set at 5000 and initial conditions at 50, with an epsilon value of 10^{-7}. Our study has a statistical power above 99% to detect the protection described for rs11209026_A in both group of patients. This power calculation has been performed with EpiInfo v5 considering allelic sample size of controls as 1092 and of patients as 790 (which is the number of MS patients, the smaller patient group), expected frequency in controls as 5.7% and odds ratio (OR) =0.26 (as described in Duerr et al.[3]).

Interaction was sought by analyzing the distribution of genotypes for one SNP in cases, conditioned by genotypes for the other SNP(s).

Results

All the polymorphisms studied conformed to Hardy–Weinberg expectations in controls. The analysis of *IL23R* polymorphisms showed susceptibility to both diseases marked by the presence of rs11209026_A (Gln) (Table 1). The other polymorphism studied, rs7517847, did not show significant differences between patients and controls.

Table 1: Genotypic and allelic frequencies of *IL23R* SNPs in CD, MS and controls.

IL23R	CD		MS		Controls	
	n	**%**	**n**	**%**	**n**	**%**
rs7517847	n=598		n=414		n=546	
TT	219	36.6	134	32.4	192	35.2
TG	279	46.7	215	51.9	260	47.6
GG	100	16.7	65	15.7	94	17.2
T	717	59.9	483	58.3	644	58.9
G	479	40.1	345	41.7	448	41.1
rs11209026	n=598		n=395		n=546	
GG	501	83.8	330	83.5	485	88.8
GA	95	15.9	59	14.9	60	11.0
AA	2	0.3	6	1.5	1	0.2
G	1097	91.7	719	91.0	1030	94.3
A	99	8.3	71	9.0	62	5.7

Carriers of rs11209026_A: p=0.01 OR=1.54 (1.08-2.20) CD vs. Controls
p=0.02 OR=1.57 (1.06-2.32) MS vs. Controls
rs11209026_A: p=0.02 OR=1.50 (1.07-2.11) CD vs. Controls
p=0.006 OR=1.64 (1.14-2.37) MS vs. Controls

Haplotype analysis does not offer additional information: the two haplotypes containing rs11209026_A behave similarly (Table 2).

Table 2: *IL23R* haplotypic frequencies in CD (n=1178), MS (n=788) and controls (n=1092).

	CD	MS	Controls
TG	58.5	56.2	58.3
GG	33.4	34.9	36.1
GA	6.63	6.93	4.95
TA	1.42	1.94	0.72

Stratification of patients by the presence of the susceptibility factors HLA-DQ2 or HLA-DRB1*1501 did not show any significant result. However, interesting results emerged when MS patients were subdivided by clinical forms (Table 3). Patients showing a PP form seemed to be more affected by *IL23R*. A significant excess of rs7517847 heterozygous individuals was observed in the group of PP patients (p=0.0003, OR=4.72, 95% CI=1.83–14.34 vs relapsing-remitting (RR) + secondary

progressive (SP) patients; p=0.0002, OR=4.77, 95% CI=1.88–14.36 vs controls). However, it is not clear whether rs11209026_A is only associated to increased susceptibility in PP patients (PP vs controls, p=0.004, OR=3.19, 95% CI=1.45–6.89) or also in RR+SP patients because borderline significant values are obtained both when RR+SP are compared to PP (p=0.04, OR=0.47, 95% CI=0.21–1.05) and to controls (p=0.04, OR=1.49, 95% CI=1.00–2.23).

Table 3: Genotypic and allelic frequencies of IL23R SNPs in MS patients stratified by clinical forms: relapsing-remitting (RR)+secondary progressive (SP) and primary progressive (PP).

IL23R	RR+SP		PP	
	n	**%**	**n**	**%**
rs7517847	n=328		n=32	
TT	116	35.4	4	12.5
TG	157	47.9	26[a]	81.3
GG	55	16.8	2	6.3
T	389	59.3	34	53.1
G	267	40.7	30	46.9
rs11209026	n=309		n=31	
GG	262	84.8	23	74.2
GA	43	13.9	6	19.4
AA	4	1.3	2	6.5
G	567	91.7	52	83.9
A	51	8.3	10[b]	16.1

[a] p=0.0003 OR=4.72 95% CI 1.83-14.34
[b] p=0.04 OR=2.14 95% CI 0.96-4.67

We also studied the possible susceptibility interaction between *IL23R* and *IL12B* polymorphisms but no effect was observed (Table 4). There are no differences in *IL12B* genotypes conditioned by *IL23R* and vice versa.

Table 4: Stratification of the IL12B data by the IL23R genotypes in CD, MS and controls.

			IL23R					
			rs7517847			**rs11209026**		
	IL12B		TT	TG	GG	GG	GA	AA
CD	**rs3212227**	AA	135	158	59	297	53	2
		AC	72	102	33	174	33	0
		CC	7	13	6	21	5	0
MS		AA	82	112	36	183	32	4
		AC	41	81	19	115	19	1
		CC	5	10	3	17	1	0
Controls		AA	105	154	57	278	37	1
		AC	73	93	31	176	21	0
		CC	11	12	6	27	2	0
CD	**rs6887695**	CC	95	114	36	208	36	1
		CG	104	127	53	235	48	1
		GG	19	36	11	55	11	0
MS		CC	64	79	31	141	23	4
		CG	60	108	28	157	30	2
		GG	7	23	3	27	4	0
Controls		CC	80	108	49	206	30	1
		CG	96	112	38	222	24	0
		GG	13	32	7	47	5	0

Discussion

We analyzed the role of *IL23R* polymorphisms in relation to CD and MS and found that the uncommon variant of the functional polymorphism Arg381Gln (rs11209026), coding for glutamine, seems to be increasing susceptibility to both inflammatory conditions. A recent study has reported lack of association of *IL23R* with CD in the Dutch population [10] and with MS in the US population [11]; this would indicate that population-specific effects may exist, as already suggested by the discordant results regarding *MYO9B* in CD [12,13] or tumor necrosis factor (*TNF*)-376A in MS [14,15]. In MS, a stronger effect was observed in the group of patients showing PP disease and further analyses are needed to clarify the involvement of Arg381Gln on patients showing RR and SP disease, which constitute the most numerous MS group. It can be argued that spurious results can emerge when very

small groups are compared but the highly significant p-value that we obtained makes this possibility unlikely. In both diseases, the Arg381Gln polymorphism seems to be associated with susceptibility independently of *IL12B* variation.

The relevance of *IL23R* polymorphisms in autoimmune diseases became first apparent from a genomewide study performed in Crohn's disease [3] and this result has been repeatedly confirmed [16–22] including in our population (manuscript submitted). Later, significant association involving this gene was also found with psoriasis and ankylosing spondylitis [4,5]. In all these studies, the allele 381Gln marks protection. However, most of the studies are being performed in relation to Crohn's disease and it seems increasingly clear that more than one association signal might be present [23,24]. The presence of two or more mutations involved in disease susceptibility would be in keeping with the results observed in our primary progressive MS patients, which showed a significant increased frequency of heterozygote individuals for the SNP rs7517847 (each allele of this SNP would be in linkage disequilibrium (LD) with a distinct etiologic mutation). The genetic region including *IL23R*, 1p31.3, shows strong linkage disequilibrium, what makes extremely difficult to attribute the risk/protection observed to any single marker. In fact, as additional SNPs are being included in the subsequent studies, the variants showing the strongest disease association are changing, which suggests that they are not themselves the etiological variants. As a consequence, *IL23R* SNPs distinct from the two studied might be causing susceptibility in the Spanish population. Notwithstanding, the missense mutation Arg381Gln displayed the strongest association in the original study [3] and association with this polymorphism was replicated in all subsequent studies, including Crohn's disease, psoriasis and ankylosing spondylitis. It is intriguing why the Arg to Gln substitution in codon 381 seems to be causing a different effect in different inflammatory conditions: protection against Crohn's disease, psoriasis and ankylosing spondylitis, and susceptibility to MS and CD. This apparent controversy could be the consequence of different factors. It could be that exactly the opposite molecular effect would be involved in Crohn's disease, and the other pathologies mentioned above, contrasting with CD and MS. Alternatively, the etiologic factor for CD and MS could be different from the etiologic variation causing protection to the other group of diseases and in this scenario both factors would be in LD with opposite alleles of the Arg381Gln polymorphism, which

by itself would act as a mere genetic marker. In this scenario, the two distinct variations could have a similar functional relevance, for example, increasing transcription but in different cell types. Although extensive *IL23R* literature is emerging, the functional role of Arg381Gln is still unknown and consequently it is not clear how to solve this issue.

A relevant role of IL-23 in MS was postulated early during IL-23 research but this statement was based mainly on murine models, specifically on studies showing that the p19 subunit and consequently IL-23 was critical in experimental autoimmune encephalomyelitis (EAE) [25], the prototypical autoimmune mouse model for MS. However, EAE and MS show an important difference regarding IFN-γ, the quintessential Th1 cytokine: IFN-γ administration leads to suppression of EAE [26] but to exacerbation of MS [27]. In addition, differences between mice and humans regarding the specific role of IL-23 are also emerging as Th17 commitment seems to be driven by IL-23 and/or IL-1β in humans [2] and by IL-6 and transforming growth factor-β in mice, in which IL-23 acts as a survival/proliferation signal for Th17 committed cells [28].

Our work evidences the involvement of *IL23R* in CD and MS, especially in the PP form. Further studies are necessary in our or another populations to confirm these results.

Acknowledgements

We are most grateful to Carmen Martínez Cuervo for her expert technical assistance. This work was supported by grants from 'Fundación Ramón Areces' and from FIS (PI070353). Concepción Núñez and Alfonso Martínez have a FIS contract (CA06/0163 and 0CP04/00175, respectively) and Elena Urcelay works for the 'Fundación para la Investigación Biomédica-Hospital Clínico San Carlos'. Bárbara Dema receives a grant from 'Fundación Mutua Madrileña'.

References

[1] Oppmann B, Lesley R, Blom B, Timans JC, Xu Y, Hunte B et al. Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with biological activities similar as well as distinct from IL-12. Immunity 2000; 13: 715–725.

[2] Wilson NJ, Boniface K, Chan JR, McKenzie BS, Blumenschein WM, Mattson JD et al. Development, cytokine profile and function of human interleukin 17-producing helper Tcells. Nat Immunol 2007; 8: 950–957.

[3] Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science 2006; 314: 1461–1463.

[4] Cargill M, Schrodi SJ, Chang M, Garcia VE, Brandon R, Callis KP et al. A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes. Am J Hum Genet 2007; 80: 273–290.

[5] Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A et al. Association scan of 14, 500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet 2007; 39: 1329–1337.

[6] Nilsen EM, Jahnsen FL, Lundin KE, Johansen FE, Fausa O, Sollid LM et al. Gluten induces an intestinal cytokine response strongly dominated by interferon gamma in patients with celiac disease. Gastroenterology 1998; 115: 551–563.

[7] Kahl KG, Kruse N, Toyka KV, Rieckmann P. Serial analysis of cytokine mRNA profiles in whole blood samples from patients with early multiple sclerosis. J Neurol Sci 2002; 200: 53–55.

[8] Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. Arch Dis Child 1990; 65: 909–911.

[9] Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. Ann Neurol 1983; 13: 227–231.

[10] Weersma RK, Zhernakova A, Nolte IM, Lefebvre C, Rioux JD, Mulder F et al. ATG16L1 and IL23R are associated with inflammatory bowel diseases but not with celiac disease in the Netherlands. Am J Gastroenterol 2007 [E-pub ahead of print].

[11] Begovich AB, Chang M, Caillier SJ, Lew D, Catanese JJ,Wang J et al. The autoimmune disease-associated IL12B and IL23R polymorphisms in multiple sclerosis. Hum Immunol 2007; 68: 934–937.
[12] Monsuur AJ, de Bakker PI, Alizadeh BZ, Zhernakova A, Bevova MR, Strengman E et al. Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. Nat Genet 2005; 37: 1341–1344.
[13] Hunt KA, Monsuur AJ, McArdle WL, Kumar PJ, Travis SP, Walters JR et al. Lack of association of MYO9B genetic variants with coeliac disease in a British cohort. Gut 2006; 55: 969–972.
[14] Fernandez-Arquero M, Arroyo R, Rubio A, Martin C, Vigil P, Conejero L et al. Primary association of a TNF gene polymorphism with susceptibility to multiple sclerosis. Neurology 1999; 53: 1361–1363.
[15] Martinez A, Rubio A, Urcelay E, Fernandez-Arquero M, De Las Heras V, Arroyo R et al. TNF-376A marks susceptibility to MS in the Spanish population: a replication study. Neurology 2004; 62: 809–810.
[16] Baldassano RN, Bradfield JP, Monos DS, Kim CE, Glessner JT, Casalunovo T et al. Association of variants of the interleukin-23 receptor gene with susceptibility to pediatric crohn's disease. Clin Gastroenterol Hepatol 2007; 5: 972–976.
[17] Borgiani P, Perricone C, Ciccacci C, Romano S, Novelli G, Biancone L et al. Interleukin-23R Arg381Gln is associated with susceptibility to Crohn's disease but not with phenotype in an Italian population. Gastroenterology 2007; 133: 1049–1051; author reply 1051–1042.
[18] Cummings JR, Ahmad T, Geremia A, Beckly J, Cooney R, Hancock L et al. Contribution of the novel inflammatory bowel disease gene IL23R to disease susceptibility and phenotype. Inflamm Bowel Dis 2007; 13: 1063–1068.
[19] Dubinsky MC, Wang D, Picornell Y, Wrobel I, Katzir L, Quiros A et al. IL-23 receptor (IL-23R) gene protects against pediatric Crohn's disease. Inflamm Bowel Dis 2007; 13: 511–515.
[20] Tremelling M, Cummings F, Fisher SA, Mansfield J, Gwilliam R, Keniry A et al. IL23R variation determines susceptibility but not disease phenotype in inflammatory bowel disease. Gastroenterology 2007; 132: 1657–1664.
[21] Van Limbergen J, Russell RK, Nimmo ER, Drummond HE, Smith L, Davies G et al. IL23R Arg381Gln is associated with childhood onset inflammatory bowel disease in Scotland. Gut 2007; 56: 1173–1174.

[22] Yamazaki K, Onouchi Y, Takazoe M, Kubo M, Nakamura Y, Hata A. Association analysis of genetic variants in IL23R, ATG16L1 and 5p13. 1 loci with Crohn's disease in Japanese patients. J Hum Genet 2007; 52: 575–583.

[23] Glas J, Seiderer J, Wetzke M, Konrad A, Torok HP, Schmechel S et al. rs1004819 is the main disease-associated IL23R variant in German Crohn's disease patients: combined analysis of IL23R, CARD15, and OCTN1/2 variants. PLoS ONE 2007; 2: e819.

[24] Raelson JV, Little RD, Ruether A, Fournier H, Paquin B, Van Eerdewegh P et al. Genome-wide association study for Crohn's disease in the Quebec founder population identifies multiple validated disease loci. Proc Natl Acad Sci USA 2007; 104: 14747–14752.

[25] Cua DJ, Sherlock J, Chen Y, Murphy CA, Joyce B, Seymour B et al. Interleukin-23 rather than interleukin-12 is the critical cytokine for autoimmune inflammation of the brain. Nature 2003; 421: 744–748.

[26] Heremans H, Dillen C, Groenen M, Martens E, Billiau A. Chronic relapsing experimental autoimmune encephalomyelitis (CREAE) in mice: enhancement by monoclonal antibodies against interferon-gamma. Eur J Immunol 1996; 26: 2393–2398.

[27] Panitch HS, Hirsch RL, Haley AS, Johnson KP. Exacerbations of multiple sclerosis in patients treated with gamma interferon. Lancet 1987; 1: 893–895.

[28] Bettelli E, Carrier Y, Gao W, Korn T, Strom TB, Oukka M et al. Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells. Nature 2006; 441: 235–238.

II. Lack of association of *NKX2-3*, *IRGM*, and *ATG16L1* inflammatory bowel disease susceptibility variants with celiac disease

Bárbara Dema[a], Miguel Fernández-Arquero[a], Carlos Maluenda[b], Isabel Polanco[c], M. Ángeles Figueredo[a], Emilio G. de la Concha[a], Elena Urcelay[a], Concepción Núñez[a]

[a] Servicio de Inmunología Clínica, Hospital Clínico San Carlos, 28040 Madrid, Spain
[b] Servicio de Pediatría, Hospital Clínico San Carlos, 28040 Madrid, Spain
[c] Servicio de Gastroenterología Pediátrica, Hospital La Paz, 28046 Madrid, Spain

Human Immunology (2009); 70(11):946-9

Resumen

Cada vez hay más evidencias de la presencia de factores de susceptibilidad comunes entre distintas enfermedades autoinmunes. Basándonos en esta idea, *NKX2-3*, *ATG16L1* e *IRGM,* que son factores de riesgo de la enfermedad inflamatoria intestinal firmemente establecidos, podrían ser nuevos genes candidatos de la enfermedad celíaca (EC). *NKX2-3* codifica un factor de transcripción que en ratones parece estar implicado en el desarrollo intestinal. Los genes *ATG16L1* e *IRGM* forman parte de la autofagia, un proceso relacionado con la inmunidad innata y adaptativa. Nuestro objetivo fue estudiar la implicación en la susceptibilidad a EC de cinco polimorfismos en estos genes: rs10883365 y rs888208 en el gen *NKX2-3,* rs2241880 en *ATG16L1*, y rs10065172 y rs4958847 en *IRGM*. Se llevaron a cabo estudios de asociación utilizando 725 pacientes celíacos españoles y 956 controles sanos étnicamente semejantes, así como también 309 tríos. Las frecuencias genéticas se compararon con el test χ^2 y el estudio familiar con el test de desequilibrio de transmisión. Las diferencias existentes entre pacientes celíacos y controles no alcanzaron valores significativos cuando se compararon las frecuencias genotípicas y alélicas. En el estudio familiar tampoco se observó una transmisión diferencial de alelos o haplotipos de padres heterocigotos a su hijo afectado. En conclusión, no existe evidencia de asociación con EC de los polimorfismos de susceptibilidad a enfermedad de Crohn, estudiados en los genes *NKX2-3*, *ATG16L1* e *IRGM*.

Abstract

Evidence about the presence of susceptibility factors shared among different autoimmune diseases is increasing. Based on this idea, *NKX2-3*, *ATG16L1,* and *IRGM* which are well-established inflammatory bowel disease risk factors, could be new celiac disease (CD) candidate genes. *NKX2-3* encodes a transcription factor that in mice seems to be involved in gut development. The *ATG16L1* and *IRGM* genes act in autophagy, a process related to innate and adaptive immunity. We aimed to study the implication of five polymorphisms in these genes in CD susceptibility: rs10883365 and rs888208 in the *NKX2-3* gene, rs2241880 in *ATG16L1*, and rs10065172 and rs4958847 in *IRGM*. Association studies were performed using 725 Spanish CD patients and 956 ethnically matched healthy controls, as well as 309 parent–child trios. Genetic frequencies were compared with the χ^2 test and the familial study used the transmission disequilibrium test. Differences between CD patients and controls did not reach significance when genotypic and allelic frequencies were compared. No differential transmission of alleles or haplotypes from heterozygous parents to affected children was observed in the familial study. In conclusion, no evidence of association with CD has been reported for the Crohn's disease susceptibility polymorphisms studied in the *NKX2-3*, *ATG16L1*, and *IRGM* genes.

KEYWORDS: celiac disease susceptibility; autophagy; single nucleotide polymorphism; *NKX2-3*.

Introduction

Celiac disease (CD) is a chronic inflammatory condition triggered by the ingestion of gluten or related proteins by genetically susceptible individuals. Important efforts have been accomplished with the aim of elucidating the different genes contributing to disease susceptibility, mainly with the development of genome-wide association studies in past years [1,2]. However, the genetic basis of CD is not completely understood and further studies are necessary to address this issue.

Crohn's disease and ulcerative colitis, the main clinical forms of inflammatory bowel disease (IBD), share persistent intestinal inflammation with CD. The occurrence of some common, genetically determined processes in the origin of these diseases has been postulated. The presence of linkage regions shared by these diseases, such as 5q31–33 (*IBD5* and *CELIAC2*) and 19p13 (*IBD6* and *CELIAC4*) [3,4], the two most promising linkage regions with the exception of the *HLA*, have long supported this idea. More recently, new clues derived from the association with ulcerative colitis and Crohn's disease of the CD susceptibility polymorphisms on the *IL2-IL21* region, at 4q27, and on the *IL18RAP* gene, at 2q12 [1,2,5,6]. Moreover, the firmly established Crohn's disease susceptibility gene *IL23R* [7] has been linked to CD in the Finnish population [8] and associated with the Spanish population [9]. Accordingly, the most recent CD susceptibility genes have been reported through an investigation of the role of polymorphisms previously associated with other autoimmune diseases, such as type 1 diabetes or IBD [10,11].

Based on this idea, polymorphisms associated with Crohn's disease and ulcerative colitis in genome-wide association studies could provide clues about candidate genes for CD, especially those supported by their function. With this in mind, we selected the *NKX2-3* (10q24) [12], *ATG16L1* (2q37) [13], and *IRGM* (5q33) [12] genes.

NKX2-3 (NK2 transcription factor related, locus 3) belongs to a family of genes that encodes transcription factors containing homeodomains and, therefore, is implicated in basic developmental functions. In mice, inactivation of the *NKX2-3* gene causes severe defects of gut development [14]. In addition, the deficient mice lack the organized T- and B-cell structure in spleen and mucosa associated lymphoid tissue

[15]. A single nucleotide polymorphism (SNP), rs10883365, located 5 kb upstream of the transcription origin, has been associated with Crohn's disease and ulcerative colitis [16–18].

The *ATG16L1* (autophagy-related 16-like 1) and the *IRGM* (immunity-related GTPase family M) genes encode proteins implicated in autophagy. Via this homeostatic cellular process, cytoplasmic portions are degraded in a regulated manner. Autophagy is also implicated in innate and adaptive immunity by delivering intracellular pathogens and other antigens, thus preventing inflammation and autoimmunity [19]. The *ATG16L1* polymorphism rs2241880 (T300A) has been extensively associated with Crohn's disease susceptibility [20]. In addition, this specific variant has been related to abnormalities in Paneth cells [21,22] and handling of intracellular pathogens or internalized antigens [23]. The expression levels of *IRGM* in several cell types control the efficiency of antibacterial autophagy [24]. The two most strongly associated SNPs in this gene were rs4958847 and rs10065172; the latter is in perfect linkage disequilibrium (r^2=1) with a copy number variation that influences the *IRGM* expression pattern [24].

In this scenario, we performed an association study with CD patients to analyze the role of several SNPs in the *NKX2-3*, *ATG16L1,* and *IRGM* genes, previously associated with IBD [12,13].

Subjects and Methods

A case–control study was carried with a maximum of 725 Spanish CD patients and 956 ethnically matched healthy individuals. Three hundred nine cases were also included, with both parents, in a familial study. Patients were diagnosed following the revised criteria of CD diagnosis proposed by the European Society for Paediatric Gastroenterology, Hepatology, and Nutrition (ESPGHAN) and collected in two centers of the same region: Hospital Clínico San Carlos and Hospital La Paz, both in Madrid. Sixty-one percent of patients were female and 92% carried the allelic susceptibility combination coding HLA-DQ2. Healthy individuals were collected from a pool of blood donors and laboratory staff. Written informed consent was obtained

from all participants. Our study was approved by the ethics committee of the above-mentioned hospitals.

We studied five SNPs by TaqMan technology (Applied Biosystems, Inc., Foster City, CA): rs10883365 (C__31657361_10) and rs888208 (C___7465608_10), located on chromosome 10 in the promoter and the 3´ untranslated region of the *NKX2-3* gene, respectively; rs2241880 (C___9095577_20), located in exon 8 of *ATG16L1*, on chromosome 2; and rs10065172 (C__30593568_10) and rs4958847 (C___1398968_10), located in the exon and immediately flanking the 3´ region of the *IRGM* gene, respectively on chromosome 5. All of these SNPs were previously associated with IBD susceptibility except rs888208 (on *NKX2-3*), which was included for haplotypic analysis. The genotyping success rate was over 95% for all assays. The lower number of samples genotyped for rs2241880 resulted from the inclusion of new patient and control samples for genotyping *NKX2-3* and *IRGM* polymorphisms.

Genotypic, allelic, and haplotypic frequencies were compared using the χ^2 test or Fisher's exact test (when expected values were below 5), implemented in the statistical package EpiInfo v.6.02 (World Health Organization, Geneva, Switzerland). Haplotypic frequencies were estimated with the expectation–maximization (EM) algorithm implemented in Haploview v.4.1 software. Linkage disequilibrium among markers of each gene was measured by calculating *D´* and r^2 values (*D´*=1, r^2=0.38, for *NKX2-3* markers and *D´*=1, r^2=0.70 for the *IRGM* studied SNPs). Family data were analyzed with the transmission disequilibrium test (TDT) to evaluate the distorted transmission of one allele or haplotype from heterozygous parents to the affected child.

Our case–control study has a statistical power of 80% calculated with EpiInfo v.6.02 to detect the following effects: rs10883365, odds ratio (OR)=1.22; rs2241880, OR=1.25; rs10065172, OR=1.44; and rs4958847, OR=1.35.

Results

All SNPs studied conformed to Hardy–Weinberg expectations.

Case–control studies were performed with the five selected polymorphisms. The genotypic and allelic frequencies did not significantly differ between CD patients and healthy controls (Table 1). The frequency of the haplotypes conformed by the SNPs studied in the *NKX2-3* and *IRGM* genes did not demonstrate statistically significant differences between both groups (data not shown). Additionally, a family study did not detect any overtransmitted allele or haplotype (where appropriate) through TDT.

No significant differences were observed after stratifying CD patients by the principal genetic susceptibility factor HLA-DQ2 (Table 2) or by gender (data not shown).

Table 1: Case-control and familial data for the SNPs studied.

SNP		Case- Control					TDT	
		CD		Controls		p value	T:U	p value
NKX2-3								
rs10883365		N=725	%	N=956	%		309 TRIOS	
	GG	189	26.1	254	26.6			
	GA	355	49.0	487	50.9			
	AA	181	25.0	215	22.5	p=0.5		
	G	733	50.6	995	52.0			
	A	717	49.4	917	48.0	p=0.4	167:156	p=0.3
rs888208		N=725	%	N=956	%		309 TRIOS	
	AA	389	53.7	527	55.1			
	AG	280	38.6	358	37.4			
	GG	56	7.70	71	7.40	p=0.8		
	A	1058	73.0	1412	73.8			
	G	392	27.0	500	26.2	p=0.6	136:107	p=0.04
ATG16L1								
rs2241880		N=607	%	N=859	%		238 TRIOS	
	GG	178	29.3	246	28.6			
	GA	314	51.7	407	47.4			
	AA	115	18.9	206	24.0	p=0.06		
	G	670	55.2	899	52.3			
	A	544	44.8	819	47.7	p=0.1	114:122	p=0.3
IRGM								
rs10065172		N=723	%	N=925	%		303 TRIOS	
	CC	553	76.5	672	72.6			
	CT	154	21.3	233	25.2			
	TT	16	2.20	20	2.20	p=0.2		
	C	1260	87.1	1577	85.2			
	T	186	12.9	273	14.8	p=0.1	71:70	p=0.5
rs4958847		N=725	%	N=925	%		302 TRIOS	
	GG	494	68.1	600	64.9			
	GA	206	28.4	287	31.0			
	AA	25	3.45	38	4.11	p=0.3		
	G	1194	82.3	1487	80.4			
	A	256	17.7	363	19.6	p=0.1	95:86	p=0.3

T= transmitted, U= untransmitted (note that only from heterozygous parents).

Table 2: SNPs studied stratified by HLA-DQ2 in celiac patients.

SNP		CD DQ2 positive		CD DQ2 negative		p value
NKX2-3						
rs10883365		N=671	%	N=53	%	
	GG	169	25.2	20	37.7	
	GA	337	50.2	17	32.1	
	AA	165	24.6	16	30.2	p=0.03
	G	675	50.3	57	53.8	
	A	667	49.7	49	46.2	p=0.5
rs888208		N=671	%	N=53	%	
	AA	357	53.2	32	60.4	
	AG	265	39.5	14	26.4	
	GG	49	7.30	7	13.2	p=0.08
	A	979	73.0	78	73.6	
	G	363	27.0	28	26.4	p=0.9
ATG16L1						
rs2241880		N=559	%	N=47	%	
	GG	164	29.3	14	29.8	
	GA	289	51.7	24	51.1	
	AA	106	19.0	9	19.1	p=0.9
	G	617	55.2	52	55.3	
	A	501	44.8	42	44.7	p=1
IRGM						
rs10065172		N=670	%	N=53	%	
	CC	510	76.1	43	81.1	
	CT	144	21.5	10	18.9	
	TT	16	2.40	0	0.00	p=0.4
	C	1164	86.9	96	90.6	
	T	176	13.1	10	9.40	p=0.3
rs4958847		N=672	%	N=53	%	
	GG	453	67.4	41	77.4	
	GA	195	29.0	11	20.8	
	AA	24	3.60	1	1.9	p=0.3
	G	1101	81.9	93	87.7	
	A	243	18.1	13	12.3	p=0.1

Discussion

In case– control studies, we analyzed whether the well established inflammatory bowel disease risk factors in the *NKX2-3*, *ATG16L1,* and *IRGM* genes are also shared by CD patients. We did not determine an association with any of the polymorphisms studied, concordant with the lack of association with T300A *ATG16L1* reported in the Dutch population [25]. Because SNPs discovered in genome-wide association studies (GWAS) might not be the causal variants but just genetic markers correlated with them, we have included haplotypic analysis to cover additional variability, but no evidence of association was detected.

GWAS are, at this moment, the most productive strategy for finding new genetic risk factors; nonetheless, it must be remembered that this high throughput often results from extending these studies through follow-up of SNPs with originally lower significant p values or by focusing on SNPs previously associated with other autoimmune diseases. In this sense, regions previously associated with CD were involved in other autoimmune diseases: *IL18RAP* in type 1 diabetes [10] or *IL2-IL21* in ulcerative colitis [5] and rheumatoid arthritis [26]. On the other hand, genes originally involved in predisposition to other diseases have now been associated with CD risk, such as *OLIG3/TNFAIP3,* which was previously associated with rheumatoid arthritis and systemic lupus erythematosus [11] and the IBD susceptibility factor *DLG5* [27]. However, the present study seems to indicate that the IBD risk factors in the *NKX2-3*, *ATG16L1,* and *IRGM* genes are not involved in CD susceptibility in the Spanish population. We chose to study these genes because they were reported to participate in IBD susceptibility. A role in CD risk could be also anticipated because of their function. However, it must be noted that attention to these three genes derived from IBD GWAS, that is, from "hypothesis-free" studies in which SNPs are included independently based on their function. Afterward, the fine mapping search for the functional implication of the observed associations begins, but the specific etiology is often hard to elucidate.

No clear relationship exists between the *NKX2-3* gene and IBD. Studies in mice relate this gene to gut development and mucosa-associated lymphoid tissue [14,15], but further investigation is necessary to establish its specific pathogenic role and

perhaps new functions more specific to IBD will be unraveled. The association of the *ATG16L1* and *IRGM* genes with Crohn's disease turned researchers' attention to autophagy as an important process in this chronic disease, but the precise link between them is still unknown. Autophagy dysregulation has been related to other pathological processes [19]; therefore, a role in CD could be also suspected. Again, additional studies in this field are warranted.

In conclusion, these well-established Crohn's disease susceptibility loci do not seem to be involved in CD susceptibility. The possibility remains that the genes studied might be relevant in CD, but the polymorphisms affecting CD predisposition would not be the ones concerning IBD risk.

Acknowledgments

We are grateful to Carmen Martínez Cuervo and M. Ángel García Martínez for their expert technical assistance. This work was supported by Project CP08/0213 from the Fondo de Investigaciones Sanitarias. Bárbara Dema receives a grant from the Fundación Mutua Madrileña. Concepción Núñez has a FIS contract (No. CP08/0213) and Elena Urcelay works for the Fundación para la Investigación Biomédica–Hospital Clínico San Carlos.

References

[1] Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M, et al. Newly identified genetic risk variants for celiac disease related to the immune response. Nat Genet 2008;40(4):395–402.
[2] van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nat Genet 2007;39:827–9.
[3] Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, McLeod RS, Griffiths AM, et al. Genome-wide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. Am J Hum Genet 2000;66:1863–70.

[4] Van Belzen MJ, Meijer JW, Sandkuijl LA, Bardoel AF, Mulder CJ, Pearson PL, et al. A major non-*HLA* locus in celiac disease maps to chromosome 19. Gastroenterology 2003;125:1032–41.

[5] Festen EA, Goyette P, Scott R, Annese V, Zhernakova A, Brant SR, et al. Genetic variants in the region harbouring IL2/IL21 associated to ulcerative colitis. Gut 2009;58:799–804.

[6] Zhernakova A, Festen EM, Franke L, Trynka G, van Diemen CC, Monsuur AJ, et al. Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring CARD9 and IL18RAP. Am J Hum Genet 2008;82:1202–10.

[7] Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science 2006;314:1461–3.

[8] Einarsdottir E, Koskinen LL, Dukes E, Kainu K, Suomela S, Lappalainen M, et al. IL23R in the Swedish, Finnish, Hungarian and Italian populations: association with IBD and psoriasis, and linkage to celiac disease. BMC Med Genet 2009; 10:8

[9] Nunez C, Dema B, Cenit MC, Polanco I, Maluenda C, Arroyo R, et al. IL23R: a susceptibility locus for celiac disease and multiple sclerosis? Genes Immun 2008;9:289–93.

[10] Smyth DJ, Plagnol V, Walker NM, Cooper JD, Downes K, Yang JH, et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. N Engl J Med 2008;359:2767–77.

[11] Trynka G, Zhernakova A, Romanos J, Franke L, Hunt K, Turner G, et al. Coeliac disease associated risk variants in TNFAIP3 and REL implicate altered NF-κB signalling. Gut 2009;58:1078–83.

[12] The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–78.

[13] Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nat Genet 2007;39:207–11.

[14] Pabst O, Zweigerdt R, Arnold HH. Targeted disruption of the homeobox transcription factor Nkx2-3 in mice results in postnatal lethality and abnormal development of small intestine and spleen. Development 1999;126:2215–25.

[15] Pabst O, Forster R, Lipp M, Engel H, Arnold HH. NKX2.3 is required for MAdCAM-1 expression and homing of lymphocytes in spleen and mucosa associated lymphoid tissue. Embo J 2000;19:2015–23.
[16] Fisher SA, Tremelling M, Anderson CA, Gwilliam R, Bumpstead S, Prescott NJ, et al. Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. Nat Genet 2008;40:710–2.
[17] Franke A, Balschun T, Karlsen TH, Hedderich J, May S, Lu T, et al. Replication of signals from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis. Nat Genet 2008;40:713–5.
[18] Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA, et al. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat Genet 2007;39:830–2.
[19] Levine B, Deretic V. Unveiling the roles of autophagy in innate and adaptive immunity. Nat Rev Immunol 2007;7:767–77.
[20] Zhang HF, Qiu LX, Chen Y, Zhu WL, Mao C, Zhu LG, et al. ATG16L1 T300A polymorphism and Crohn's disease susceptibility: evidence from 13,022 cases and 17,532 controls. Hum Genet 2009;125:627–31.
[21] Cadwell K, Liu JY, Brown SL, Miyoshi H, Loh J, Lennerz JK, et al. A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells. Nature 2008;456:259–63.
[22] Cadwell K, Patel KK, Komatsu M, Virgin HW 4th, Stappenbeck TS. A common role for Atg16L1, Atg5 and Atg7 in small intestinal Paneth cells and Crohn disease. Autophagy 2009;5:250–2.
[23] Kuballa P, Huett A, Rioux JD, Daly MJ, Xavier RJ. Impaired autophagy of an intracellular pathogen induced by a Crohn's disease associated ATG16L1 variant. PLoS ONE 2008;3:e3391
[24] McCarroll SA, Huett A, Kuballa P, Chilewski SD, Landry A, Goyette P, et al. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. Nat Genet 2008;40:1107–12.
[25] Weersma RK, Zhernakova A, Nolte IM, Lefebvre C, Rioux JD, Mulder F, et al. ATG16L1 and IL23R are associated with inflammatory bowel diseases but not with celiac disease in the Netherlands. Am J Gastroenterol 2008;103:621–7.
[26] Teixeira VH, Pierlot C, Migliorini P, Balsa A, Westhovens R, Barrera P, et al. Testing for the association of the KIAA1109/Tenr/IL2/IL21 gene region with

rheumatoid arthritis in a European family-based study. Arthritis Res Ther 2009;11:R45

[27] Festen EA, Zhernakova A, Wijmenga C, Weersma RK. Association of DLG5 variants with gluten-sensitive enteropathy. Gut 2008;57:1027–8.

III. Autoimmune disease association signals in *CIITA* and *KIAA0350* are not involved in celiac disease susceptibility

Bárbara Dema[a], Alfonso Martínez[a], Miguel Fernández-Arquero[a], Carlos Maluenda[b], Isabel Polanco[c], M. Ángeles Figueredo[a], Emilio G. de la Concha[a], Elena Urcelay[a], Concepción Núñez[a]

[a] Servicio de Inmunología Clínica, Hospital Clínico San Carlos, 28040 Madrid, Spain
[b] Servicio de Pediatría, Hospital Clínico San Carlos, 28040 Madrid, Spain
[c] Servicio de Gastroenterología Pediátrica, Hospital La Paz, 28046 Madrid, Spain

Tissue Antigens (2009) 73, 326-329

Resumen

La enfermedad celíaca (EC) es una enfermedad multifactorial caracterizada por una inflamación intestinal tras la exposición al gluten en individuos genéticamente susceptibles. La influencia de ciertos alelos de los genes *HLA* (*human leukocyte antigen*) (aquellos que codifican los heterodímeros HLA-DQ2 y DQ8) está firmemente establecida, pero esto no explica el riesgo genético global de la enfermedad. *CIITA* puede ser un buen gen candidato a la susceptibilidad de EC debido a que se encuentra principalmente regulado transcripcionalmente y que codifica el principal regulador de la transcripción génica de las moléculas del complejo principal de histocompatibilidad tipo II. *CIITA* se encuentra localizado en 16p13, región que contiene también el gen *KIAA0350* (*CLEC16A*), asociado con dos enfermedades autoinmunes en estudios de barrido genómico (GWAS, *genome-wide association studies*). Nuestro objetivo es estudiar la implicación de polimorfismos en *CIITA* y *KIAA0350* en la susceptibilidad a EC, con especial atención en la evaluación de la posible presencia de más de un factor de riesgo en la región. Realizamos un estudio caso-control con 607 pacientes celíacos y hasta 779 controles sanos, todos españoles. Todas las muestras fueron genotipadas para cinco polimorfismos de un único nucleótido: rs3087456 (2168A/G) y rs4774 en *CIITA* and rs7203459, rs6498169 y rs2903692 en *KIAA0350*. No se obtuvieron resultados estadísticamente significativos al comparar las frecuencias genotípicas, alélicas y haplotípicas entre enfermos y controles. Nuestros resultados parecen descartar la influencia de los marcadores en *CIITA* y *KIAA0350* previamente asociados con otras enfermedades autoinmunes, en la susceptibilidad a EC.

Abstract

Celiac disease (CD) is a multifactorial disease characterized by intestinal inflammation after gluten exposure in genetically susceptible individuals. A strong influence of certain human leukocyte antigen (*HLA*) alleles (those coding the HLA-DQ2 and DQ8 heterodimers) is well established, but they cannot explain the overall genetic risk. *CIITA* could be a good candidate gene for CD because it is mainly transcriptionally regulated, and it encodes the master regulator of major histocompatibilty complex class II gene transcription. *CIITA* is located in 16p13, a region also containing *KIAA0350* (*CLEC16A*), associated with two autoimmune diseases in genome-wide association studies. We aimed at studying the involvement of polymorphisms in *CIITA* and *KIAA0350* in CD susceptibility, with special attention to evaluate the possible presence of more than one risk factor in the region. We performed a case–control study with 607 CD patients and up to 779 healthy controls, all Spaniards. All samples were genotyped for five single nucleotide polymorphisms: rs3087456 (2168A/G) and rs4774 in *CIITA* and rs7203459, rs6498169 and rs2903692 in *KIAA0350*. No significant results were obtained when comparing genotypic, allelic or haplotypic frequencies between patients and controls. Our results seem to discard the influence in CD susceptibility of *CIITA* and *KIAA0350* markers previously associated with other autoimmune diseases.

KEYWORDS: 16p13; celiac disease susceptibility; *CIITA*; *CLEC16A*.

Introduction

Celiac disease (CD) is a complex disease with prevalence around 1% in populations of European ancestry. Presentation of gluten-derived peptides by human leukocyte antigen (HLA)-DQ2 molecules seems to be essential for development of CD, as evidenced by the high proportion of CD patients (90%–95%) carrying the HLA alleles coding the DQ2 heterodimer. However, those alleles are present in approximately 30% of healthy people, and therefore, additional risk factors, genetic or environmental, are required. The relevance of the HLA-DQ2 heterodimer in CD stresses the possibility that modulation of the expression of the HLA genes may be also contributing to CD susceptibility. The expression of HLA class II genes (including DQ genes) is regulated mainly at the transcriptional level by different factors that bind highly conserved elements present in the promoters of those genes and serve as a scaffold for the non-DNA-binding class II transactivator C2TA [1], which holds the complex together. Because *CIITA* is the only of these regulatory components with a complex pattern of expression, which indeed parallels that of the major histocompatibilty complex (MHC) class II genes, it seems to be the master transcriptional regulator of those MHC genes [2]. Expression of *CIITA* is itself regulated mainly by transcription, and the single nucleotide polymorphism (SNP) rs3087456 (2168A/G), located in the CIITA promoter region, has been involved in the differential transcription of MHC molecules after interferon-γ stimulation. Moreover, this polymorphism has been associated with two autoimmune diseases with an inflammatory component as rheumatoid arthritis and multiple sclerosis [3], although controversial results exist [4–6]. It is already known that some susceptibility loci are shared between CD and other autoimmune diseases. As a matter of fact, the recently discovered CD susceptibility region containing the IL2/IL21 genes [7] has been also associated with type 1 diabetes, rheumatoid arthritis and psoriatic arthritis [8, 9].

The *CIITA* gene is located in 16p13, and interestingly, polymorphisms in this genetic region have been associated to type 1 diabetes and multiple sclerosis in the recent genome-wide association studies (GWAS). In these diseases, the strongest association signal has been found with SNPs located in *KIAA0350* (*CLEC16A*, C-type lectin domain family 16, member A). This gene encodes a protein that is predicted to be a C-type lectin. This kind of proteins is characterized by the presence of domains

that recognize carbohydrates, usually pathogen associated. Because bacteria involvement has been postulated in CD aetiology [10], *KIAA0350* could be also a candidate gene for CD susceptibility. In addition, the linkage disequilibrium (LD) existing between *CIITA* and *KIAA0350* makes the inclusion of these two loci mandatory when planning the study of *CIITA*: the close proximity of two susceptibility regions could preclude detection of true associations [11].

Although the lack of association with the 16p13 region reported in the CD GWAS [7, 12] could discourage us to perform this study, it should be kept in mind that GWAS analyze the isolated effect of different SNPs. However, other kind of genetic variation could be underlying the pathological process, and individual SNPs must show a high correlation with the causal variant to allow the detection of a significant signal, as happened with the deletion upstream the *IRGM* locus involved in Crohn's disease, that showed perfect disequilibrium (r^2=1) with one SNP included in the GWAS [13]. However, this co-occurrence may not be very common and haplotype analysis (not included in GWAS) may cover more genetic variation than individual SNPs. In addition, the negative result concerning 16p13 obtained in the GWAS could be a consequence of its stringent threshold for significance and a modest association could be still present. It is also relevant that a different population has been analyzed in the present study (CD GWAS was performed in the UK population).

In this work, we aimed at studying the implication of *CIITA* and *KIAA0350* polymorphisms in CD susceptibility, with special care in avoiding the possible masking effect of one locus over the other.

Material and methods

Samples from 607 CD patients and up to 779 healthy controls, all white Spaniards, were collected in two centers of the same region (Hospital Clínico San Carlos and Hospital La Paz, Madrid) after obtaining the corresponding written informed consent. Controls are mainly blood donors and hospital employees. CD was diagnosed according to the European Society for Paediatric Gastroenterology and Nutrition criteria [14]. Ninety-four per cent of our patients were diagnosed below 18

years, 61% were female and 92% HLA-DQ2 positive. Ethical approval for the study was obtained from the Ethics Committee of the Hospital Clínico San Carlos.

Two SNPs were studied in the *CIITA* gene: rs3087456 (2168A/G), located in the promoter, and rs4774 (nt1614 from coding sequence), in exon 12 and causing an amino acid replacement from glycine to alanine. These two SNPs conformed haplotypes shown to capture the susceptibility/protective signal to several autoimmune diseases better than each of those markers separately [15]. In *KIAA0350*, we studied three SNPs: rs7203459, rs6498169 and rs2903692. The last two SNPs were selected because of their known association with multiple sclerosis and type 1 diabetes, respectively, in their corresponding GWAS. The first SNP was selected for haplotypic analysis.

All SNPs were genotyped by TaqMan technology following conditions recommended by the manufacturer (Applied Biosystems Inc., Foster City, CA). Success genotyping score was above 96% for all the SNPs studied in cases and controls.

Genotypic, allelic and haplotypic frequencies were compared between patients and controls using the chi-square test. Haplotypes were estimated with the expectation–maximization algorithm implemented in the software HAPLOVIEW (and their frequencies were very similar to those directly calculated from data obtained from 225 families). Statistical power calculations were performed using EPIINFO v5 software.

The possible existence of population stratification was assessed with the analysis of five European Ancestry Informative Markers: rs12440216, rs12905017, rs2060983, rs1404402 and rs1805005. No significant differences between cases and controls were detected [$p>0.05$, odds ratios (ORs) between 1.03 and 1.18].

All the polymorphisms studied showed consistency with Hardy–Weinberg equilibrium.

Results

Linkage disequilibrium values measured with the *D´* and r^2 coefficients between all the markers studied are shown in Table 1.

Table 1: Linkage disequilibrium values between the five SNPs studied.

	rs3087456	rs4774	rs7203459	rs2903692	rs6498169
rs3087456	-	0.24	0.02	0.04	0.03
rs4774	0	-	0.38	0.22	0
rs7203459	0	0.01	-	0.94	0.85
rs2903692	0	0.01	0.45	-	0.99
rs6498169	0	0	0.12	0.32	-

Data calculated from our control sample.
D' values are shown above diagonal and r^2 lower diagonal.

No significant differences between CD patients and controls were observed when each marker was analyzed separately, when considering either *CIITA* or *KIAA0350* polymorphism (Table 2). Association was not detected either after haplotype analysis (Table 3). Because some degree of linkage disequilibrium is present between some markers located in these two different loci, extended haplotypes including the five studied SNPs were also included in the analysis. The five SNP haplotypes let us perform stratified analysis to avoid obtaining negative results as a consequence of the presence of different susceptibility factors in linkage disequilibrium with different alleles of the same marker. No significant effects were observed.

Stratified analysis to investigate the differential involvement of the markers studied in CD patients positive and negative for HLA-DQ2 were also carried out, but no significant differences were observed (data not shown).

Table 2: Genotypic and allelic frequencies in CD patients and controls.

Gene	CD		Controls		p, OR (95% CI)
CIITA					
rs3087456	n=600	%	n=519	%	
AA	336	56.0	296	57.0	0.72
AG	221	36.8	192	37.0	
GG	43	7.2	31	6.0	
A	893	74.4	784	75.5	0.54,
G	307	25.6	254	24.5	OR=1.06 (0.87-1.29)
rs4774	n=598	%	n=562	%	
GG	325	54.3	316	56.2	0.78
GC	233	39.0	212	37.7	
CC	40	6.7	34	6.0	
G	883	73.8	844	75.1	0.48,
C	313	26.2	280	24.9	OR=0.94 (0.77-1.13)
KIAA0350					
rs7203459	n=606	%	n=686	%	
TT	336	55.4	368	53.6	0.71
TC	235	38.8	272	39.7	
CC	35	5.8	46	6.7	
T	907	74.8	1008	73.5	0.43,
C	305	25.2	364	26.5	OR=0.93 (0.78-1.12)
rs2903692	n=607	%	n=779	%	
GG	203	33.4	277	35.6	0.21
GA	316	52.1	370	47.5	
AA	88	14.5	132	16.9	
G	722	59.5	924	59.3	0.93,
A	492	40.5	634	40.7	OR=0.99 (0.85-1.16)
rs6498169	n=607	%	n=548	%	
AA	270	44.5	258	48.5	0.28
AG	278	45.8	227	42.7	
GG	59	9.7	63	11.8	
A	818	67.4	743	67.8	0.83,
G	396	32.6	353	32.2	OR=1.02 (0.85-1.22)

p values are calculated for the 3*2 table (genotypic comparison) and for the 2*2 table (allelic comparison).
OR values are referred to the minor allele.

Table 3: Haplotypic frequencies for C*IITA* (rs3087456-rs4774) and *KIAA0350* (rs7203459-rs2903692- rs6498169) in CD patients and controls.

Haplotypes		CD patients		Controls		p, OR (95% CI)
CIITA		n=1200	%	n=1130	%	
AG		634	52.8	620	54.9	0.32, OR=0.92 (0.78-1.09)
AC		259	21.6	232	20.5	0.53, OR=1.07 (0.87-1.31)
GG		253	21.1	228	20.2	0.59 OR=1.06 (0.86-1.30)
GC		54	4.5	50	4.4	0.93 OR=1.02 (0.67-1.54)
KIAA0350		n=1213	%	n=1380	%	
TGG		387	31.9	438	31.7	0.93 OR=1.01 (0.85-1.19)
TGA		327	26.9	370	26.8	0.93 OR=1.01 (0.84-1.20)
CAA		295	24.3	351	25.4	0.51 OR=0.94 (0.78-1.13)
TAA		195	16.1	209	15.1	0.51 OR=1.07 (0.86-1.33)
CIITA KIAA0350		n=1195	%	n=1075	%	
AG	TGG	203	16.8	187	17.1	0.81
AG	CAA	161	13.3	164	15.0	0.23
AG	TGA	152	12.5	141	12.9	0.79
AG	TAA	118	9.7	103	9.4	0.80
AC	TGA	92	7.6	78	7.2	0.70
AC	TGG	91	7.5	68	6.2	0.22
AC	CAA	40	3.3	38	3.5	0.85
AC	TAA	37	3.0	37	3.4	0.59
GG	CAA	91	7.5	72	6.6	0.39
GG	TGG	68	5.6	67	6.1	0.64
GG	TGA	56	4.6	48	4.4	0.80
GG	TAA	37	3.0	30	2.7	0.67
GC	TGA	26	2.1	22	2.0	0.80
GC	TGG	23	1.9	20	1.8	0.81

Only haplotypes with frequencies higher than 1% are shown.

Discussion

The recent GWAS have yielded some new genetic regions implicated in CD [7, 12]. However, the genetic basis of this disease is far to be completely understood. In this work, we have investigated the role of different polymorphisms previously associated with other autoimmune diseases [3, 15–17] and located in two genes in the chromosomal region 16p13, *CIITA* and *KIAA0350*. We did not observe any significant result when allele or haplotype frequencies were compared between CD patients and

controls despite having above 80% of statistical power to detect an allelic effect (expressed as OR) between 1.3 and 1.4, depending on the SNP considered. With haplotype analyses, we tried to cover additional association signals not marked by the individual SNPs as well as to evaluate possible masking signals emerging from different susceptibility effects located in the same region.

CIITA had been associated with two autoimmune diseases [3, 15], although this effect is not free of controversy [4–6]. Moreover, the recently discovered association of *KIAA0350* polymorphisms in the GWAS of type 1 diabetes and multiple sclerosis suggested the possibility that the previous *CIITA* associations were because of the influence of *KIAA0350*, although whether those associations are independent is difficult to be discerned because of the linkage disequilibrium existent between polymorphisms in those two genes. However, none of the genetic polymorphisms shown to be useful susceptibility markers in other autoimmune diseases is valid in CD. Specifically, the lack of influence of *CIITA* in CD risk suggests that the strong association existent between CD and MHC is not somehow regulated by *CIITA*, although the possibility remains that other genetic variants not included in this work are relevant.

In conclusion, this work seems to indicate the lack of involvement in CD of the polymorphisms in *CIITA* and *KIAA0350* previously described as relevant genetic risk factors in other autoimmune conditions.

Acknowledgments

We are most grateful to Carmen Martínez Cuervo and M. Ángel García Martínez for their expert technical assistance. This work was supported by a grant from ‘Fundación Ramón Areces’ and FIS 070353. Bárbara Dema receives a grant from ‘Fundación Mutua Madrileña’. Concepción Núñez and Alfonso Martínez have a FIS contract (CA06/0163 and CP04/00175, respectively) and Elena Urcelay works for the ‘Fundación para la Investigación Biomédica-Hospital Clínico San Carlos’.

References

[1] Ting JP, Trowsdale J. Genetic control of MHC class II expression. Cell 2002: 109 (Suppl): S21–33.

[2] Leibund Gut-Landmann S, Waldburger JM, Krawczyk M et al. Mini-review: specificity and expression of CIITA, the master regulator of MHC class II genes. Eur J Immunol 2004:34: 1513–25.

[3] Swanberg M, Lidman O, Padyukov L et al. MHC2TA is associated with differential MHC molecule expression and susceptibility to rheumatoid arthritis, multiple sclerosis and myocardial infarction. Nat Genet 2005: 37: 486–94.

[4] Akkad DA, Jagiello P, Szyld P et al. Promoter polymorphism rs3087456 in the MHC class II transactivator gene is not associated with susceptibility for selected autoimmune diseases in German patient groups. Int J Immunogenet 2006: 33: 59–61.

[5] Bronson PG, Criswell LA, Barcellos LF.The MHC2TA-168A/G polymorphism and risk for rheumatoid arthritis: a meta-analysis of 6861 patients and 9270 controls reveals no evidence for association. Ann Rheum Dis 2008: 67: 933–6.

[6] Sartoris S, Brendolan A, Degola A et al. Analysis of CIITA encoding AIR-1 gene promoters in insulin-dependent diabetes mellitus and rheumatoid arthritis patients from the northeast of Italy: absence of sequence variability. Hum Immunol 2000:61: 599–604.

[7] van Heel DA, Franke L, Hunt KA et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nat Genet 2007: 39: 827–9.

[8] Liu Y, Helms C, Liao W et al. A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. PLoS Genet 2008: 4: e1000041.

[9] Zhernakova A, Alizadeh BZ, Bevova M et al. Novel association in chromosome 4q27 region with rheumatoid arthritis and confirmation of type 1 diabetes point to a general risk locus for autoimmune diseases. Am J Hum Genet 2007: 81: 1284–8.

[10] Forsberg G, Fahlgren A, Horstedt P, Hammarstrom S, Hernell O, Hammarstrom ML. Presence of bacteria and innate immunity of intestinal epithelium in childhood celiac disease. Am J Gastroenterol 2004: 99: 894–904.

[11] Nunez C, Santiago JL, Varade J et al. IL4 in the 5q31 context: association studies of type 1 diabetes and rheumatoid arthritis in the Spanish population. Immunogenetics 2008: 60: 19–23.
[12] Hunt KA, Zhernakova A, Turner G et al. Newly identified genetic risk variants for celiac disease related to the immune response. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nat Genet 2008: 40: 395–402.
[13] McCarroll SA, Huett A, Kuballa P et al. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. Nat Genet 2008: 40:1107–12.
[14] Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. Arch Dis Child 1990: 65: 909–11.
[15] Martinez A, Sanchez-Lopez M, Varade J et al. Role of the MHC2TA gene in autoimmune diseases. Ann Rheum Dis 2007: 66: 325–9.
[16] Hafler DA, Compston A, Sawcer S et al. Risk alleles for multiple sclerosis identified by a genome-wide study. N Engl J Med 2007: 357: 851–62.
[17] Hakonarson H, Grant SF, Bradfield JP et al. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. Nature 2007: 448: 591–4.

3. - ESTUDIO DE GENES CANDIDATOS

I. The *IL6-174G/C* polymorphism is associated with celiac disease susceptibility in girls.

II. Study of the promoter of the *NOS2A* gene in celiac disease risk.

I. The *IL6* -174G/C polymorphism is associated with celiac disease susceptibility in girls

Bárbara Dema[a], Alfonso Martínez[a], Miguel Fernández-Arquero[a], Carlos Maluenda[b], Isabel Polanco[c], M. Ángeles Figueredo[a], Emilio G. de la Concha[a], Elena Urcelay[a], Concepción Núñez[a]

[a] Servicio de Inmunología Clínica, Hospital Clínico San Carlos, 28040 Madrid, Spain
[b] Servicio de Pediatría, Hospital Clínico San Carlos, 28040 Madrid, Spain
[c] Servicio de Gastroenterología Pediátrica, Hospital La Paz, 28046 Madrid, Spain

Human Immunology (2009) 70, 191-194

Resumen

El objetivo de este artículo fue estudiar el papel de polimorfismos en los genes *IL6* e *IL6R* en la susceptibilidad a enfermedad celíaca (EC). Debido a que en la literatura se describen diferencias en género relacionadas con la IL-6, se llevó a cabo un análisis estratificado por sexo. Realizamos un estudio caso-control con 333 pacientes celíacos pediátricos y 853 controles sanos, todos ellos españoles, y un estudio familiar con fines de replicación con una muestra independiente que incluye 303 tríos. Se genotiparon tres polimorfismos de un único polimorfismo que cubren la mayor parte de la variabilidad del gen *IL6* (rs2069827, rs1800795 [-174G/C] y rs2069840) y un polimorfismo funcional en el gen *IL6R* (rs8192284, Asp358Ala), utilizando tecnología TaqMan. Las comparaciones del estudio caso-control fueron llevadas a cabo con el test χ^2 y el estudio familiar se analizó con el test de desequilibrio de transmisión. No se detectó asociación con la EC global, de cualquiera de los polimorfismos testados. Sin embargo tras estratificación por sexo, se observó que la variante del promotor *IL6* -174C aumenta el riesgo a desarrollar EC específicamente en pacientes mujeres. Este efecto se muestra tanto en el estudio caso-control como en el estudio familiar (considerando las niñas incluidas en ambos estudios vs niños: p=0.021, odds ratio [OR]=1.31, intervalo de confianza [CI] al 95% 1.03–1.66; and vs controles: p=0.003, OR=1.30, 95% CI 1.09–1.55). El polimorfismo funcional *IL6* -174G/C parece influir en la susceptibilidad de niñas celíacas. El papel específico de IL-6 en el género en esta patología debe ser investigado más a fondo.

Abstract

The aim of this paper was to study the role of *IL6* and *IL6R* polymorphisms in celiac disease (CD) susceptibility. Because previous literature describes IL-6-related gender differences, sex-stratified analyses were performed. We undertook a case–control study with 333 pediatric CD patients and 853 healthy controls, all white Spaniards, and a family study using an independent sample including 303 trios for replication purposes. Three single-nucleotide polymorphisms tagging most of the variability of the *IL6* gene (rs2069827, rs1800795 [-174G/C], and rs2069840) and one functional polymorphism in *IL6R* (rs8192284, Asp358Ala) were genotyped using TaqMan technology. Case–control comparisons were performed with the χ^2 test and family data were analyzed with the transmission disequilibrium test. No association was observed between any tested polymorphism and overall CD. However, after sex stratification, we found that the *IL6* promoter variant -174C increases the risk of developing CD specifically in female patients. This effect was observed both in the case–control and in the family studies (considering girls included in both studies vs boys: p=0.021, odds ratio [OR]=1.31, 95% confidence interval [CI] 1.03–1.66; and vs controls: p=0.003, OR=1.30, 95% CI 1.09–1.55). The functional -174G/C *IL6* polymorphism seems to influence CD susceptibility in girls. The gender-specific role of IL-6 in this pathology must be further investigated.

KEYWORDS: interleukin-6; celiac disease susceptibility; single-nucleotide polymorphism; girls.

Introduction

Celiac disease (CD) is a chronic inflammatory disorder of special relevance in children because it is the most common form of food hypersensitivity in this population group, with a prevalence of around 1 per 100 children in European populations [1]. Patients with CD may display different symptoms, but its classical form is characterized by malabsorption and failure to gain weight caused by gluten ingestion in genetically susceptible individuals. The strongest genetic susceptibility factor is located in the *HLA* region (alleles coding the heterodimers HLA-DQ2 or HLA-DQ8), as observed in most autoimmune diseases. Recently published genome wide association studies have offered very few additional CD susceptibility loci, adding only around 4% to the already known *HLA* heritability, estimated to be around 40% [2,3]. For this reason, it is necessary to continue the search for genetic susceptibility factors following different approaches (for instance, analyzing selected subsets of patients).

Interleukin-6 (IL-6) is a cytokine with several important functions in the immune system because it is involved in innate and adaptive responses [4]. IL-6 has been extensively studied in the context of several inflammatory or immune conditions and different results have been obtained. In CD, increased levels of IL-6 have been observed in the serum of untreated patients [5] or after "*in vitro*" stimulation of CD biopsies with gluten [6]. Therefore, genetic polymorphisms modifying IL-6 levels could potentially contribute to CD susceptibility. However, few and relatively underpowered studies have been performed with this aim [7,8] and additional analysis are still necessary. We will evaluate the putative role of genetic *IL6* polymorphisms in pediatric CD risk by performing a single-nucleotide polymorphism (SNP) tagging of the gene, with special attention to include the promoter -174G/C variation, which has been extensively studied because of its involvement in IL-6 transcriptional activity and plasma levels [9]. In addition, we will study a functional polymorphism in the *IL6R* gene, Ala358Asp, located in the proteolytic cleavage site of IL-6Rα that consequently alters the soluble IL-6Rα receptor (sIL-6Rα) levels [10]. This receptor form seems to be essential in the regulation of IL-6 functions [11–13].

The different prevalence between males and females observed in most autoimmune diseases could be partly the result of sex related differences in certain proinflammatory cytokines. In this regard, IL-6 production has been investigated and gender differences have been detected [14,15]. We will also evaluate the possible gender-specific susceptibility conferred by the genetic polymorphisms studied.

Subjects and Methods

Subjects

We studied 333 white Spanish celiac disease patients and 853 ethnically matched healthy controls (mostly blood donors and laboratory staff). Additionally, we analyzed an independent replication sample comprising 303 trios constituting affected individuals and progenitors. All samples were consecutively collected from two centers in the same region (Hospital Clínico San Carlos and Hospital La Paz, Madrid). CD was diagnosed following the European Society for Paediatric Gastroenterology and Nutrition criteria [16]. All patients included in the present study were diagnosed before the age of 18 years, 61% were female, and 97% were HLA-DQ2 and/or HLA-DQ8 positive. Written informed consent was obtained from all participants. Ethical approval for the study was obtained from the Ethics Committee of the Hospital Clínico San Carlos.

Genotyping

DNA was extracted from fresh peripheral blood leukocytes by a "salting out" procedure. We analyzed three SNPs in *IL6* (Figure 1), on chromosome 7, which tag over 80% of the variability of this gene: rs2069827, rs1800795, and rs2069840. This tagging was performed using the "aggressive tagging" option present in the Haploview software, with r^2 threshold set at 0.8 and minimum minor allele frequency at 0.1. We forced the inclusion of the classical *IL6* promoter polymorphism -174G/C (rs1800795). The three selected SNPs are in a linkage disequilibrium block of $D' = 1$ and their correlation values measured by the r^2 coefficient are 0.097 between rs2069827 and rs1800795, 0.051 between rs2069827 and rs2069840, and 0.476

between rs1800795 and rs2069840 (data obtained from http://www.hapmap.org). Additionally, we studied the *IL6R* functional polymorphism on 1q21, rs8192284 (Asp358Ala) [10]. All polymorphisms were genotyped using TaqMan genotyping assays under conditions recommended by the manufacturer (Applied Biosystems, Inc., Foster City, CA, USA). Genotype call rate success was over 97% for the three SNPs studied in all sample groups.

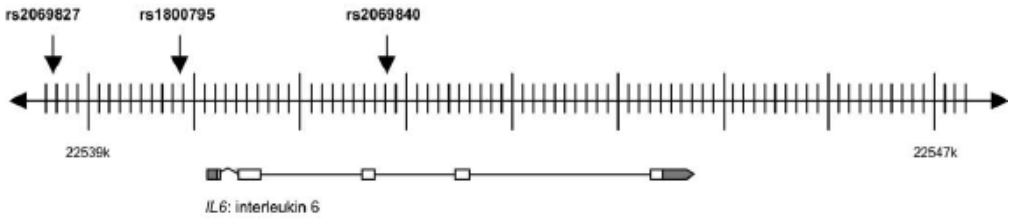

Figure 1: Location of the three *IL6* studied polymorphisms.

Statistical analysis

Genotype, allele, and haplotype frequencies were compared between cases and controls by applying χ^2 tests to the corresponding 3 x 2, 2 x 2 or 2 x *n* contingency table, respectively. Family data were analyzed using the transmission disequilibrium test (TDT), which only uses information provided by heterozygous parents to assess the preferential transmission of one allele or haplotype over the other(s). Haplotypic frequencies were either estimated using the expectation–maximization algorithm implemented in the Haploview program [17] for the samples included in the case–control study or deduced directly from the pedigree for the patients used in the family study. Our case–control study has 80% statistical power to detect an odds ratio (OR)=1.7 for rs2069827 and an OR=1.3 for the other three studied polymorphisms; statistical power was calculated using the statistical package EpiInfo v5.00 (CDC, Atlanta, GA, USA).

Interaction between *IL6* and *IL6R* polymorphisms was sought by analyzing the distribution of genotypes in cases for each *IL6* SNP, conditioned by genotypes for the *IL6R* SNP.

The possible existence of population stratification was assessed with the analysis of five European Ancestry Informative markers: rs12440216, rs12905017, rs2060983, rs1404402, and rs1805005. No significant differences between cases and controls were detected ($p>0.05$, OR between 1.03 and 1.18).

Results

All SNPs studied conformed to Hardy–Weinberg proportions.

Genotypic and allelic frequencies in CD patients and controls are shown in Table 1. Significant differences between both groups were observed when rs2069827 and rs8192284 were analyzed (allelic test: $p=0.02$, OR=1.53, 95% confidence interval [CI] 1.07–2.20 and $p=0.02$, OR=1.24, 95% CI 1.03–1.49, respectively). In concordance with previous studies, we also performed these analyses after stratification by gender (Table 2). No significant differences were observed between female and male allelic frequencies in controls; therefore, they were considered together to increase statistical power. The -174C (rs1800795) allele was more common in female CD patients when compared with males ($p=0.07$, OR=1.35, 95% CI 0.96–1.91) or with controls ($p=0.02$, OR=1.31, 95% CI 1.04–1.64). Note the increased statistical power to detect this effect in comparison with controls because of the higher sample size of this group. Therefore, a lower *p* value is obtained, but similar ORs emerge from both comparisons. Male CD patients and controls show a similar frequency of the -174C allele ($p=0.80$, OR=0.96, 95% CI 0.72–1.29). A significant difference after sex stratification was also observed with rs2069840 ($p=0.009$), but in this case it is not as obvious which group differs from controls.

Table 1: Genotypic and allelic frequencies in CD patients and controls, and TDT results.

SNP	CD		Controls		Case-control	Trios	
	n	**%**	**n**	**%**	**p, OR (95%CI)**	**T:U**	**p**
IL6							
rs2069827	n=333		n=848			n=302	
GG	278	83.5	759	89.5	0.01		
GT	55	16.5	84	9.9			
TT	0	0	5	0.6			
G	611	91.7	1602	94.5			
T	55	8.26	94	5.5	0.02, 1.53 (1.07-2.20)	39:43	0.37
rs1800795	n=332		n=835			n=303	
GG	136	41.0	386	46.2	0.26		
GC	154	46.4	358	42.9			
CC	42	12.7	91	10.9			
G	426	64.2	1130	67.7			
C	238	35.8	540	32.3	0.10, 1.17 (0.96-1.42)	155:135	0.13
rs2069840	n=333		n=842			n=302	
CC	143	42.9	348	41.3	0.84		
CG	147	44.1	389	46.2			
GG	43	12.9	105	12.5			
C	433	65.0	1085	64.4			
G	233	35.0	599	35.6	0.79, 0.97 (0.80-1.18)	133:139	0.38
IL6R							
rs8192284	n=332		n=838			n=294	
AA	113	34.0	325	38.8	0.04		
AC	155	46.7	398	47.5			
CC	64	19.3	115	13.7			
A	381	57.4	1048	62.5			
C	283	42.6	628	37.5	0.02, 1.24 (1.03-1.49)	120:137	0.16

n indicates the number of individuals or the number of trios, in the case-control or the familial analyses, respectively. ORs are referred to the minor allele. T=transmitted; U=untransmitted (note that only from heterozygous parents).

Table 2: Genotypic and allelic frequencies in CD patients after gender stratification, and TDT results in female and male offspring.

SNP	CD women		CD men		Women vs. men	Women		Men	
	n	%	n	%	p, OR (95%CI)	T:U	p	T:U	p
IL6									
rs2069827	n=203		n=131			n=190		n=112	
GG	170	83.7	109	83.2	0.90				
GT	33	16.3	22	16.8					
TT	0	0	0	0					
G	373	91.9	240	91.6					
T	33	8.13	22	8.40	0.90, 0.97 (0.53-1.76)	23:29	0.24	16:14	0.43
rs1800795	n=203		n=130			n=190		n=113	
GG	78	38.4	59	45.4	0.15				
GC	94	46.3	60	46.2					
CC	31	15.3	11	8.46					
G	250	61.6	178	68.5					
C	156	38.4	82	31.5	0.07, 1.35 (0.96-1.91)	104:80	0.045	51:55	0.39
rs2069840	n=203		n=131			n=188		n=114	
CC	101	49.7	42	32.1	0.005				
CG	77	37.9	70	53.4					
GG	25	12.3	19	14.5					
C	279	68.7	154	58.8					
G	127	31.3	108	41.2	0.009, 0.65 (0.46-0.91)	76:93	0.109	57:46	0.162
IL6R									
rs8192284	n=203		n=130			n=185		n=109	
AA	65	32.0	48	36.9	0.45				
AC	95	46.8	61	46.9					
CC	43	21.2	21	16.2					
A	225	55.4	157	60.4					
C	181	44.6	103	39.6	0.21, 1.23 (0.88-1.70)	77:83	0.35	43:54	0.16

n indicates the number of individuals or the number of trios, in the case-control and in the familial analyses, respectively. T=transmitted; U=untransmitted (note that only from heterozygous parents). ORs are referred to the minor allele.

Haplotypic analysis with the *IL6* polymorphisms does not offer additional information (data not shown). Significant results were only observed when considering the haplotypes equivalent to SNPs previously showing a significant result in the individual analysis.

The analysis of the *IL6R* SNP, Asp358Ala, did not result in significant differences between females and males (p=0.21, OR=1.23, 95% CI 0.88–1.70).

Because borderline significant results were obtained in most of the case–control comparisons, we tried to replicate these associations in a familial study with independent CD patients. The results of these analyses are shown in the last columns of Tables 1 and 2. The associations with disease overall observed with *IL6* rs2069827 and *IL6R* rs8192284 were not confirmed by the familial analysis performed using the TDT (p=0.37 and p=0.16, respectively). An overtransmission of the -174C allele was also observed only when female offspring were considered, both when compared with the random distribution (vs 50% T, p=0.045) and when compared with the distribution observed in males (vs 48% T, p=0.014). For rs2069840, only a trend was observed in the TDT (vs 50% T, p=0.109), but a significant difference emerged when compared with the transmission frequency observed in boys (vs 55% T, p=0.004). No association with gender of the *IL6R* polymorphism studied could be observed with the TDT, in keeping with the results of the previous case–control study.

To increase the statistical power to evaluate the CD effect size, we pooled all CD patients (included in the case–control and in the familial studies) to perform a new case–control study: the -174C allele increases the risk specifically to females with an OR = 1.30, 95% CI 1.09–1.55, and p=0.003; and the rs2069840 major allele (C) with an OR = 1.22, 95% CI 1.02–1.46, and p=0.028 (comparisons are calculated vs controls).

We did not detect an interaction between any *IL6* polymorphism and the *IL6R* SNP studied. We also analyzed the possible interaction between -174G/C and the *IL6R* polymorphism only in the subgroup of girls, but no interaction emerged.

Discussion

We have performed a case–control study to investigate the role of IL-6 on CD susceptibility and a familial study including trios with one affected child. Our results indicate that the presence of the minor allele of the functional promoter polymorphism -174G/C (rs1800795) increases the risk of CD to females. This association observed in the case–control study was subsequently confirmed with the TDT. Moreover, with the family study we wanted to discard the chance of obtaining a positive association as a consequence of population stratification, although this seemed unlikely when European Ancestry Informative markers were analyzed. Some association may also exist with rs2069840, although our results are not conclusive. However, it must be noted that -174G/C and rs2069840 demonstrate the highest correlation of all pairs considered in this work ($r^2 = 0.26$ in our sample) and, therefore, they can demonstrate association as a consequence of the same etiologic factor (the two alleles that seem to be causing susceptibility are more frequently found together in the population, *i.e.*, they are in positive linkage disequilibrium). Despite our results, further studies are necessary to confirm the observed associations in other populations and ethnic groups.

The *IL6* locus has not been associated with CD in the genome wide scans performed [2, 3], but our study evidenced that gender stratification is required to detect the association described here. Unfortunately, no information with regard to gender stratification is available online and we cannot perform a meta-analysis with those data for further confirmation. The previous negative result from a study in the Finnish population using 106 families may raise a similar concern [8]. In agreement with the present data, although under a different autoimmune condition, such as type 1 diabetes, a positive association was observed in a TDT study in the Danish population: 90 of 137 informative transmissions to female offspring were -174C alleles [18], whereas no distorted transmission to male patients was observed.

As indicated in the Introduction, a gender bias may exist in IL-6 production [14]. In addition, gender differences observed in several pathological processes as hepatocellular carcinoma, response after traumatic injury, and hemorrhagic shock or myocardial infarction have been related to IL-6 [19–21]. A differential IL-6 production as a consequence of oestrogen modulation has been also proposed [15].

However, it is likely that several combined processes underlie the IL-6 differences observed between genders.

The finding of a significant association of *IL6* with female CD patients increases the list of relevant cytokines, from a genetic point of view, in this pathology. IL-6 exhibits important and diverse functions in immune and inflammatory processes; recently, the newly discovered subset of T helper 17 (Th17) cells has expanded the IL-6 functions. Th17 cells protect against certain extracellular pathogens, but during pathological processes they seem to be involved in inflammatory and autoimmune conditions. Differentiation of naive Th cells into effector cells depends on the cytokine microenvironment. It seems that IL-6 may be important in Th-17 commitment, because it has been described as upregulating the IL-23R expression and promoting Th17 differentiation through synergism with IL-23 [22]. Based on murine models, a role of IL-6 in connection with its relevance in Th17 differentiation has been suggested for other autoimmune diseases such as rheumatoid arthritis and multiple sclerosis [23]. However, the involvement of IL-6 in CD could be the result of several other factors. The cytokine stimulates the growth of plasma cells and the synthesis and secretion of antibodies; therefore, it could contribute to the production of IgA or IgG autoantibodies observed in most CD patients. Further studies analyzing the IL-6 role on CD are needed.

Acknowledgments

We are grateful to Carmen Martínez Cuervo and M. Ángel García Martínez for their expert technical assistance. This work was supported by a grant from "Fundación Ramón Areces," which was not involved in any way in the elaboration of the manuscript. Bárbara Dema received a grant from "Fundación Mutua Madrileña." Concepción Núñez and Alfonso Martínez have a FIS contract (CA06/0163 and CP04/00175, respectively) and Elena Urcelay works for the "Fundación para la Investigación Biomédica-Hospital Clínico San Carlos."

References

[1] Maki M, Mustalahti K, Kokkonen J, Kulmala P, Haapalahti M, Karttunen T, et al. Prevalence of celiac disease among children in Finland.NEngl J Med 2003;348: 2517–24.

[2] van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nat Genet 2007;39:827–9.

[3] Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M, et al. Newly identified genetic risk variants for celiac disease related to the immune response. Nat Genet 2008;40:395–402.

[4] Ishihara K, Hirano T. IL-6 in autoimmune disease and chronic inflammatory proliferative disease. Cytokine Growth Factor Rev 2002;13:357–68.

[5] Romaldini CC, Barbieri D, Okay TS, Raiz R Jr, Cancado EL. Serum soluble interleukin-2 receptor, interleukin-6, and tumor necrosis factor-alpha levels in children with celiac disease: response to treatment. J Pediatr Gastroenterol Nutr 2002;35:513–7.

[6] Nilsen EM, Jahnsen FL, Lundin KE, Johansen FE, Fausa O, Sollid LM, et al. Gluten induces an intestinal cytokine response strongly dominated by interferon gamma in patients with celiac disease. Gastroenterology 1998;115:551–63.

[7] Garrote JA, Arranz E, Gomez-Gonzalez E, Leon AJ, Farre C, Calvo C, et al. IL6, IL10 and TGFB1 gene polymorphisms in coeliac disease: differences between DQ2 positive and negative patients. Allergol Immunopathol (Madr) 2005;33:245–9.

[8] Woolley N, Mustalahti K, Maki M, Partanen J. Cytokine gene polymorphisms and genetic association with coeliac disease in the Finnish population. Scand J Immunol 2005;61:51–6.

[9] Fishman D, Faulds G, Jeffery R, Mohamed-Ali V, Yudkin JS, Humphries S, et al. The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. J Clin Invest 1998;102:1369–76.

[10] Mullberg J, Oberthur W, Lottspeich F, Mehl E, Dittrich E, Graeve L, et al. The soluble human IL-6 receptor. Mutational characterization of the proteolytic cleavage site. J Immunol 1994;152:4958–68.

[11] Jones SA, Horiuchi S, Topley N, Yamamoto N, Fuller GM. The soluble interleukin 6 receptor: mechanisms of production and implications in disease. FASEB J 2001;15:43–58.
[12] Qi L, Rifai N, Hu FB. Interleukin-6 receptor gene variations, plasma interleukin-6 levels, and type 2 diabetes in U.S. Women. Diabetes 2007;56: 3075–81.
[13] Rabe B, Chalaris A, May U, Waetzig GH, Seegert D, Williams AS, et al. Transgenic blockade of interleukin 6 trans-signaling abrogates inflammation. Blood 2008; 111:1021–8.
[14] O'Connor MF, Motivala SJ, Valladares EM, Olmstead R, Irwin MR. Sex differences in monocyte expression of IL-6: role of autonomic mechanisms. Am J Physiol Regul Integr Comp Physiol 2007;293:R145–51.
[15] Prieto J. Inflammation, HCC and sex: IL-6 in the centre of the triangle. J Hepatol 2008;48:380–1.
[16] Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. Arch Dis Child 1990;65:909–11.
[17] Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263–5.
[18] Kristiansen OP, Nolsoe RL, Larsen L, Gjesing AM, Johannesen J, Larsen ZM, et al. Association of a functional 17 beta-estradiol sensitive IL6-174G/C promoter polymorphism with early-onset type 1 diabetes in females. Hum Mol Genet 2003;12:1101–10.
[19] Naugler WE, Sakurai T, Kim S, Maeda S, Kim K, Elsharkawy AM, et al. Gender disparity in liver cancer due to sex differences in MyD88-dependent IL-6 production. Science 2007;317:121–4.
[20] Samnegard A, Hulthe J, Silveira A, Ericsson CG, Hamsten A, Eriksson P. Gender specific associations between matrix metalloproteinases and inflammatory markers in post myocardial infarction patients. Atherosclerosis 2009;202: 550–6.
[21] Sperry JL, Friese RS, Frankel HL, West MA, Cuschieri J, Moore EE, et al. Male gender is associated with excessive IL-6 expression following severe injury. J Trauma 2008;64:572–8.
[22] Yang XO, Panopoulos AD, Nurieva R, Chang SH, Wang D, Watowich SS, et al. STAT3 regulates cytokine-mediated generation of inflammatory helper T cells. J Biol Chem 2007;282:9358–63.

[23] Ogura H, Murakami M, Okuyama Y, Tsuruoka M, Kitabayashi C, Kanamoto M, et al. Interleukin-17 promotes autoimmunity by triggering a positive-feedback loop via interleukin-6 induction. Immunity 2008;29:628–36.

II. Study of the promoter of the *NOS2A* gene in celiac disease risk

Bárbara Dema[1], Miguel Fernández-Arquero[1], Isabel Polanco[2], Carlos Maluenda[3], M. Ángeles Figueredo[1], Emilio G. de la Concha[1], Elena Urcelay[1], Concepción Núñez[1]

[1] Servicio de Inmunología Clínica, Hospital Clínico San Carlos, 28040 Madrid, Spain
[2] Servicio de Gastroenterología Pediátrica, Hospital La Paz, 28046 Madrid, Spain
[3] Servicio de Pediatría, Hospital Clínico San Carlos, 28040 Madrid, Spain

Resumen

Pacientes con enfermedad celíaca (EC) muestran unos niveles elevados de óxido nítrico como consecuencia de una mayor expresión del gen que codifica la óxido nítrico sintasa inducible (gen *NOS2A*), que se regula principalmente a nivel transcripcional. Nuestro objetivo en este estudio fue investigar la relevancia en EC de polimorfismos genéticos en el promotor de *NOS2A*, algunos de ellos previamente descritos como implicados en una mayor expresión de iNOS. Llevamos a cabo un estudio caso-control con 379 pacientes con EC y 702 individuos sanos, y un estudio familiar que incluye 264 tríos. Las muestras fueron genotipadas para dos polimorfismos de un único nucleótido (SNPs) (rs2779251 y rs2779248) y dos microsatélites (el pentanucleótido $(CCTTT)_n$ (rs3833912) y la inserción/deleción TAAA (rs12720460)). Los SNPs fueron genotipados utilizando tecnología TaqMan, y los microsatélites mediante PCR seguida de electroforesis capilar. Las comparaciones del estudio caso-control se realizaron con tests Chi-cuadrado o el test exacto de Fisher. El Test de Desequilibrio de Transmisión (TDT) se utilizó para analizar los datos de familias. Los haplotipos se estimaron utilizando el algoritmo EM. El estudio caso-control y el análisis con TDT no muestran ningún resultado significativo bien al analizar los marcadores individualmente o bien al combinarlos en haplotipos. Por tanto, la susceptibilidad a EC no parece verse afectada por la presencia de polimorfismos específicos en el promotor del gen *NOS2A*.

Abstract

Celiac disease (CD) patients show high levels of nitric oxide as a consequence of a higher expression of the gene coding the inducible nitric oxide synthase (*NOS2A* gene), that is mainly regulated at transcriptional level. Our aim in this study was to assess the relevance in CD of genetic polymorphisms in the promoter of *NOS2A*, some of them previously described as being implicated in a greater expression of iNOS. We performed a case-control study with 379 CD patients and 702 healthy individuals, and a family study including 264 trios. Samples were genotyped for two single nucleotide polymorphisms (SNPs) (rs2779251 and rs2779248) and two microsatellites (pentanucleotide $(CCTTT)_n$ (rs3833912) and the insertion/deletion TAAA (rs12720460)). The SNPs were genotyped using TaqMan technology, and the microsatellites by PCR followed by capillary electrophoresis. Case-control comparisons were performed by Chi-Square tests or the Fisher's exact test. Transmission disequilibrium test (TDT) was used to analyse familial data. Haplotypes were estimated using the EM algorithm. Case-control comparisons and TDT analyses did not show any significant result when markers were analysed individually or combined in haplotypes. Therefore, CD susceptibility does not seem to be affected by the presence of specific promoter *NOS2A* polymorphisms.

KEYWORDS: *NOS2A*; iNOS; celiac disease susceptibility; single-nucleotide polymorphism; microsatellite.

Introduction

Celiac disease is a chronic inflammatory bowel disease triggered by ingested gluten or related proteins in genetically predisposed individuals. Genes in the Human Leukocyte Antigen (*HLA*) region are the most important genetic factors known to be involved in CD (98% of CD patients are HLA-DQ2 and/or HLA-DQ8 positive). Last years, with the arrival of the genome wide association studies (GWAS) era, genetic components involved in CD risk outside HLA were also found [126, 132]. However, the risk attributed to HLA [43] and the new loci identified by GWAS, altogether, only explain nearly 54% of the CD genetic component [51, 157]. Therefore, new CD contributing genes remain to be discovered.

An increase in nitric oxide (NO) levels has been observed in duodenal enterocytes from CD patients. Three different isoforms of the nitric oxide synthase (NOS) have been characterized: two constitutive (endothelial and neuronal) and one inducible (iNOS) [184]. It seems that the increased NO levels found in CD are generated by the iNOS isoform as a consequence of increased expression of its gene (*NOS2A*) in the small intestine [185]. Accordingly to this, iNOS has been postulated to be involved in CD pathogenesis. This idea is supported by studies showing a decrease of nitric oxide end-products in plasma of CD patients after introduction of a gluten-free diet [102] and by in vitro studies in which gluten-gliadin induces iNOS expression in IFN-γ stimulated murine macrophages [186].

Nitric oxide production depends mainly on iNOS expression, which is regulated mostly at transcriptional level, through several proinflammatory cytokines as IFN-γ, IL-1β and TNF-α [187]. The *NOS2A* gene is located on chromosome 17 and contains several polymorphisms, some of them with a functional role, mainly those at the promoter site [188]. Some *NOS2A* polymorphisms have been previously associated with the pathogenesis of several immune diseases as rheumatoid arthritis, systemic lupus erythematosus or multiple sclerosis [162, 189, 190], but many discrepancies occur on this respect.

Therefore, our aim in this work was to study the involvement of iNOS on CD, analyzing several *NOS2A* promoter polymorphisms, possibly involved in the regulation of *NOS2A* transcription.

Materials and Methods

We performed a case-control study comprising 379 CD patients and 702 healthy individuals, and a family study with 264 trios. All individuals have Spanish ancestry and were collected in the Clínico San Carlos Hospital and La Paz Hospital, both from Madrid, Spain. CD patients were diagnosed following the ESPHANG criteria (European Society for Paediatric Gastroenterology, Hepatology and Nutrition) [173]. Sixty one per cent of these patients were female and 92% carried the main genetic susceptibility factor HLA-DQ2. Most controls were blood donors and laboratory staff. A written informed consent was signed by all individuals and the study was approved by the Ethics Committee of the Clínico San Carlos Hospital.

We genotyped four polymorphisms located at the promoter of the *NOS2A* gene (17q11.1): two microsatellite polymorphisms clasically studied: a pentanucleotide tandem repeat $(CCTTT)_n$ (rs3833912) and the insertion/deletion TAAA (4 or 3 repeats, respectively) (rs12720460) at 2.6kb and 0.7kb upstream of the transcription origin, respectively; and two single nucleotide polymorphisms (SNPs): rs2779251 and rs2779248, flanking the microsatellites (Figure 1).

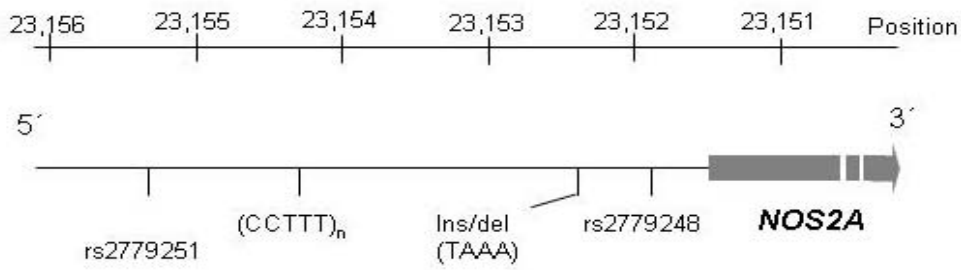

Figure 1: Location of the polymorphisms studied in the promoter of the NOS2A gene.

The SNPs were genotyped using TaqMan technology (Applied Biosystems, Foster City, CA, USA) following conditions suggested by the manufacturer. The microsatellite markers were typed by PCR with labelled primers followed by electrophoresis on an ABI 3100 sequencer. The forward primer used for rs3833912 was 5'-Fam-ACCCCTGGAAGCCTACAACT-3' and the reverse primer 5'-GCCACTGCACCCTAGCCTGTCTCA-3'; for rs12720460, we used the forward primer 5'-Hex-TGGTGCATGCCTGTAGTCC-3' and the reverse primer 5'-GAGGCCTCTGAGATGTTGGTC-3'.

Haplotypes were estimated using the EM (Expectation-Maximization) algorithm implemented in the Haploview v.4.1 software (bi-allelic markers) or the Arlequin v2.000 software (when the pentanucleotide microsatellite was included). Linkage disequilibrium (LD) among markers was measured by calculating D' and r2 values.

Genotypic, allelic and haplotypic comparisons between patients and controls were performed with the chi-square test or the Fisher´s exact test (when expected values are below 5), offered by the statistical package EpiInfo v.6.02 (World Health Organization, Geneva, Switzerland).

Family results were analysed using the transmission disequilibrium test (TDT), which compares the differential transmission of alleles or haplotypes from heterozygous parents to their affected child.

Results

All the polymorphisms studied conformed to Hardy-Weinberg expectations.

Genotypic and allelic frequencies of the polymorphisms studied did not differ among patients and controls (Table 1 and 2). The familial study did not show differences in transmission either, except for the $(CCTTT)_{12}$ allele, that was over transmitted, and $(CCTTT)_{10}$ and $(CCTTT)_{14}$ alleles that were under transmitted (Table 2). However, we did not observe this effect in the case-control study and TDT signification did not withstand multiple testing correction.

Table 1: Genotypic and allelic frequencies of the NOS2A bi-allelic markers, in celiac disease (CD) patients, controls and family analysis.

NOS2A	CD		Controls		Case-control	TDT	
rs2779251	N=379	%	N=702	%	p, OR (95%CI)	264 trios	p
CC	268	70.7	519	73.9	0.3		
CT	98	25.8	168	23.9			
TT	13	3.40	15	2.10			
C	634	83.6	1206	85.9			
T	124	16.3	198	14.1	0.16, 1.19 (0.93-1.53)	85T:70U	0.13
rs12720460 (ins/delTAAA)	N=379	%	N=702	%			
del/del	293	77.3	531	75.6	0.6		
del/ins	76	20.1	157	22.4			
ins/ins	10	2.60	14	1.90			
del	662	87.3	1219	86.8			
ins	96	12.6	185	13.2	0.7, 0.96 (0.73-1.25)	49T:68U	0.05
rs2779248	N=379	%	N=702	%			
AA	149	39.3	296	42.2	0.3		
AG	166	43.8	309	44.0			
GG	64	16.9	97	13.8			
A	464	61.2	901	64.2			
G	294	38.8	503	35.8	0.17, 1.13 (0.94-1.37)	123T:123U	0.5

N indicates the number of individuals in the case-control analysis. ORs are referred to the minor allele. T=transmitted; U=untransmitted (note that only from heterozygous parents).

Table 2: Allelic frequencies of $(CCTTT)_n$, in celiac disease (CD) patients and controls and TDT results.

NOS2A $(CCTTT)_n$		CD		Controls		Case-control	TDT		
allele	size	2N=752	%	2N=1380	%	p	T	U	p
7	170	1	0.13	3	0.22	0.5	1	1	0.7
8	175	17	2.26	31	2.25	0.9	12	17	0.2
9	180	38	5.05	70	5.07	0.9	28	27	0.5
10	185	100	13.3	189	13.7	0.8	40	59	0.03
11	190	140	18.6	259	18.8	0.9	91	79	0.2
12	195	238	31.6	430	31.2	0.8	116	89	0.03
13	200	128	17.0	230	16.7	0.8	72	74	0.5
14	205	61	8.11	117	8.48	0.8	30	46	0.04
15	210	22	2.90	37	2.68	0.7	15	11	0.3
16	215	7	0.93	12	0.87	0.9	4	7	0.9
17	220	0	0	1	0.07	0.6	2	1	0.5
18	225	0	0	1	0.07	0.6	1	1	0.7

N indicates the number of individuals in the case-control analysis. ORs are referred to the minor allele. T=transmitted; U=untransmitted (note that only from heterozygous parents).

When haplotypes conformed by the bi-allelic markers were estimated using the EM algorithm, a low number of haplotypes emerged (Table 3). This was due to the strong linkage disequilibrium (LD) among those markers (D'=0.97-1.00).

Table 3: Haplotypic frequencies with bi-allelic markers (rs2779251-rs12720460-rs2779248) in celiac disease (CD) patients and controls, and TDT results.

NOS2A haplotypes	CD		Controls		Case-control	TDT		
	2N=758	%	2N=1404	%	p, OR (95%IC)	T	U	p
C-del-A	464	61.2	897	63.9	0.2, 0.89 (0.74-1.07)	123	123	0.52
T-del-G	124	16.4	195	13.9	0.12, 1.21 (0.94-1.56)	85	70	0.13
C-ins-G	96	12.7	178	12.7	0.9, 1.0 (0.76-1.31)	49	68	0.05
C-del-G	74	9.80	127	9.00	0.6, 1.09 (0.8-1.49)	54	50	0.38

N indicates the number of individuals in the case-control analysis. T=transmitted; U=untransmitted.

When haplotypes with all the studied polymorphisms were estimated, it is observed that the pentanucleotide microsatellite breaks that disequilibrium pattern, probably as a consequence of its high mutation rate. Despite the lower LD existing with the $(CCTTT)_n$ microsatellite, a comprehensive analysis of the promoter haplotypes showed that LD values between the different alleles of this microsatellite and the other promoter markers covered a wide range of values and certain

microsatellite alleles were not randomly distributed, appearing preferentially on specific haplotypes (Table 4). Nonetheless, no statistical differences in the frequency of these haplotypes were found when comparing patients and controls.

Table 4: Linkage disequilibrium (LD) among all the studied promoter markers.

Bi-allelic promoter haplotype	(CCTTT)n	LD	
		D'	p-value
C-del-A	10	0.6	$1.3*10^{-10}$
T-del-G	8	0.4	$5.3*10^{-10}$
C-ins-G	14	0.3	$5.3*10^{-18}$
C-ins-G	13	0.3	$9.5*10^{-21}$
C-del-G	11	0.3	$1.8*10^{-10}$

When the polymorphisms studied were stratified by the main CD genetic risk factor, HLA-DQ2, or by gender, we did not find any statistical difference.

Discussion

In this study, we performed a case-control and a familial study including several polymorphisms at the promoter of the *NOS2A* gene to investigate their possible role in celiac disease susceptibility.

No statistically significant results were observed regarding any SNP or microsatellite analyzed. Moreover, haplotype analysis were performed to cover most of the variability in that promoter region but association with CD was not observed either. The implication in CD of the *NOS2A* gene was previously studied in an independent Spanish sample [191], also with negative results, althought their sample was not high enough to offer a conclusive result and only multiallelic $(CCTTT)_n$ microsatellite was involved.

The *NOS2A* promoter has been repeatedly studied and it has been reported to be one of the largest and more complex human promoters [192]. Different genetic regulatory elements have been described. For example, the tetranucleotide microsatellite $(TAAA)_n$ has been related to the basal expression of iNOS and the pentanucleotide

microsatellite $(CCTTT)_n$ to the inducible expression [188]. However, our study suggest that *NOS2A* promoter polymorphisms are not underlying the changes in NO levels described in CD patients and related to CD risk. Nevertheless, important regulatory elements of cytokine inducibility seem to be also present in regions up to 16kb upstream of the transcription site. According to this, other regulatory mechanisms are implicated in *NOS2A* expression, and they may be affecting CD susceptibility. Moreover, the study of the promoter variability does not exclude a possible implication in CD of coding (or other) variants in the *NOS2A* gene.

Althought it might be considered that the *NOS2A* region was not detected in CD susceptibility by GWAS and therefore it should not be focus of study, GWAS do not cover all the genetic variability as rare variants and structural variants, and more extensive analysis could offer interesting results.

This work evidences that variation at the promoter location of *NOS2A* gene does not affect to CD susceptibility. More exhaustive analysis would be necessary to discard the implication of this gene in CD risk.

Acknowledgements

We are most grateful to Carmen Martínez Cuervo and M. Ángel García Martínez for her expert technical assistance.

Concepción Núñez has a FIS contract (CP08/0213) and Elena Urcelay works for the "Fundación para la Investigación Biomédica-Hospital Clínico San Carlos".

References

[1] Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M, Romanos J, Dinesen LC, Ryan AW, Panesar D, Gwilliam R, Takeuchi F, McLaren WM, Holmes GK, Howdle PD, Walters JR, Sanders DS, Playford RJ, Trynka G, Mulder CJ, Mearin ML, Verbeek WH, Trimble V, Stevens FM, O'Morain C, Kennedy NP, Kelleher D, Pennington DJ, Strachan DP, McArdle WL, Mein CA, Wapenaar MC, Deloukas P, McGinnis R, McManus R, Wijmenga C, van Heel DA: Newly identified genetic risk variants for celiac disease related to the immune response. Nat Genet 40(4):395, 2008.

[2] van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, Wapenaar MC, Barnardo MC, Bethel G, Holmes GK, Feighery C, Jewell D, Kelleher D, Kumar P, Travis S, Walters JR, Sanders DS, Howdle P, Swift J, Playford RJ, McLaren WM, Mearin ML, Mulder CJ, McManus R, McGinnis R, Cardon LR, Deloukas P, Wijmenga C: A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nat Genet 39(7):827, 2007.

[3] Green PH, Jabri B: Celiac disease. Annu Rev Med 57:207, 2006.

[4] Romanos J, van Diemen CC, Nolte IM, Trynka G, Zhernakova A, Fu J, Bardella MT, Barisani D, McManus R, van Heel DA, Wijmenga C: Analysis of HLA and non-HLA alleles can identify individuals at high risk for celiac disease. Gastroenterology 137(3):834, 2009.

[5] Sollid LM: Coeliac disease: dissecting a complex inflammatory disorder. Nat Rev Immunol 2(9):647, 2002.

[6] Alderton WK, Cooper CE, Knowles RG: Nitric oxide synthases: structure, function and inhibition. Biochem J 357(Pt 3):593, 2001.

[7] Murray IA, Daniels I, Coupland K, Smith JA, Long RG: Increased activity and expression of iNOS in human duodenal enterocytes from patients with celiac disease. Am J Physiol Gastrointest Liver Physiol 283(2):G319, 2002.

[8] Spencer HL, Daniels I, Shortland J, Long RG, Murray IA: Effect of a gluten-free diet on plasma nitric oxide products in coeliac disease. Scand J Gastroenterol 39(10):941, 2004.

[9] De Stefano D, Maiuri MC, Iovine B, Ialenti A, Bevilacqua MA, Carnuccio R: The role of NF-kappaB, IRF-1, and STAT-1alpha transcription factors in the iNOS gene

induction by gliadin and IFN-gamma in RAW 264.7 macrophages. J Mol Med 84(1):65, 2006.
[10] Geller DA, Billiar TR: Molecular biology of nitric oxide synthases. Cancer Metastasis Rev 17(1):7, 1998.
[11] Kleinert H, Schwarz PM, Forstermann U: Regulation of the expression of inducible nitric oxide synthase. Biol Chem 384(10-11):1343, 2003.
[12] Barcellos LF, Begovich AB, Reynolds RL, Caillier SJ, Brassat D, Schmidt S, Grams SE, Walker K, Steiner LL, Cree BA, Stillman A, Lincoln RR, Pericak-Vance MA, Haines JL, Erlich HA, Hauser SL, Oksenberg JR: Linkage and association with the NOS2A locus on chromosome 17q11 in multiple sclerosis. Ann Neurol 55(6):793, 2004.
[13] Gonzalez-Gay MA, Llorca J, Sanchez E, Lopez-Nevot MA, Amoli MM, Garcia-Porrua C, Ollier WE, Martin J: Inducible but not endothelial nitric oxide synthase polymorphism is associated with susceptibility to rheumatoid arthritis in northwest Spain. Rheumatology (Oxford) 43(9):1182, 2004.
[14] Oates JC, Levesque MC, Hobbs MR, Smith EG, Molano ID, Page GP, Hill BS, Weinberg JB, Cooper GS, Gilkeson GS: Nitric oxide synthase 2 promoter polymorphisms and systemic lupus erythematosus in african-americans. J Rheumatol 30(1):60, 2003.
[15] Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. Arch Dis Child 65(8):909, 1990.
[16] Rueda B, Lopez-Nevot MA, Pascual M, Ortega E, Maldonado J, Lopez ML, Koeleman BP, Martin J: Polymorphism of the inducible nitric oxide synthase gene in celiac disease. Hum Immunol 63(11):1062, 2002.
[17] Kroncke KD, Fehsel K, Kolb-Bachofen V: Inducible nitric oxide synthase in human diseases. Clin Exp Immunol 113(2):147, 1998.

4.- RATIFICACIÓN DE SEÑALES SUGERENTES DE SUSCEPTIBILIDAD

I. Multiple common variants for celiac disease influencing immune gene expression.

I. Multiple common variants for celiac disease influencing immune gene expression

Patrick CA Dubois[1,37], Gosia Trynka[2,37], Lude Franke[1,2],Karen A Hunt[1], Jihane Romanos[2], Alessandra Curtotti[3], Alexandra Zhernakova[4], Róza Ádány[5], Arpo Aromaa[6], Maria Terasa Bardella[7,8], Leonard H van den Berg[9], Nicholas A Bockett[1], Emilio G. de la Concha[10], Bárbara Dema[10], Rudolf S Fehrmann[2], Miguel Fernández-Arquero[10], Szilvia Fiatal[5,11], Elvira Grandone[12], Peter M Green[13], Harry JM Groen[14], Rhian Gwilliam[15],Graham AR Heap[1], Roderick HJ Houwen[16], Sarah E Hunt[15], Katri Kaukinen[17], Dermot Kelleher[18], Ilma Korponay-Szabo[19, 20], Kalle Kurppa[17], Padraic MacMathuna[21], Markku Mäki[17], Maria Cristina Mazzilli[22], Owen T McCann[15], M Luisa Mearin[23], Charles A Mein[3], Muddassar M Mirza[13], Vanisha Mistry[1], Barbara Mora[22], Katherine I Morley[15], Chris J Mulder[24], Joseph A Murray[25], Concepción Núñez[10], Elvira Oosterom[2], Roel A Ophoff[26, 27], Isabel Polanco[28], Leena Peltonen[15, 29], Mathieu Platteel[2], Anna Rybak[30], Veikko Salomaa[6], Joachim J Schweizer[23], Maria Pia Sperandeo[31], Greetje J Tack[24], Graham Turner[18], Wieke HM Verbeek[24], Rinse K Weersma[2,32], Victorien M Wolters[16], Elena Urcelay[10], Bozena Cukrowska[33], Luigi Greco[31], Susan L. Neuhausen[34], Ross McManus[18], Donatella Barisani[35], Panos Deloukas[15], Jeffrey C Barrett[15], Paivi Saavalainen[36], Cisca Wijmenga[2], David A van Heel[1]

[1]Blizard Institute of Cell and Molecular Science, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London E1 2AT, UK
[2]Genetics Department, University Medical Center and Groningen University, 9700 RB Groningen, The Netherlands
[3]The Genome Centre, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK
[4]Division of Biomedical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands
[5]Department of Preventive Medicine, University of Debrecen, Debrecen, Hungary
[6]National Institute for Health and Welfare, Helsinki 00271, Finland
[7]Fondazione IRCCS Ospedale Maggiore Policlinico, Mangiagalli e Regina Elena, Milan, Italy
[8]Department of Medical Sciences, University of Milan, Italy
[9]Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Centre Utrecht, Heidelbergaan 100, Utrecht, The Netherlands
[10]Clinical Immunology Department, Hospital Clínico San Carlos, Madrid, Spain

[11]Public Health Research Group of Hungarian Academy of Sciences, Medical & Health Science Center, University of Debrecen, Debrecen, Hungary
[12]Unita' di Aterosclerosi e Trombosi, I.R.C.C.S Casa Sollievo della Sofferenza , S. Giovanni Rotondo, Foggia, Italy
[13]NIHR GSTFT/KCL Comprehensive Biomedical Research Centre King's College London School of Medicine 7th Floor Tower Wing Guy's Hospital London SE19RT United Kingdom
[14]Dept. of Pulmonology, University Medical Center and Groningen University, 9700 RB Groningen, The Netherlands
[15]Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK
[16]Department of Paediatric Gastroenterology, University Medical Centre Utrecht, Utrecht, The Netherlands
[17]Paediatric Research Centre, University of Tampere Medical School and Tampere University Hospital, Tampere, Finland
[18]Department of Clinical Medicine, Institute of Molecular Medicine, Trinity College Dublin, Dublin, Ireland
[19]Heim Pal Childrens Hospital, Budapest, Hungary
[20]Department of Pediatrics, Medical and Health Science Center, University of Debrecen, Hungary
[21]Gastrointestinal Unit, Mater Misericordiae University Hospital, Eccles St., Dublin 7, Ireland
[22]Department of Experimental Medicine, Sapienza University of Rome, Rome, Italy
[23]Department of Paediatrics, Leiden University Medical Centre, Leiden, The Netherlands
[24]Department of Gastroenterology, VU Medical Center, 1007 MB Amsterdam, The Netherlands
[25]Division of Gastroenterology and Hepatology, Department of Medicine, Mayo Clinic College of Medicine, Rochester, MN 55905, USA
[26]Department of Medical Genetics and Rudolf Magnus Institute, University Medical Center Utrecht, Utrecht, The Netherlands
[27]Center for Neurobehavioral Genetics, University of California, Los Angeles, CA, USA
[28]Pediatric Gastroenterology Department, Hospital La Paz, Madrid, Spain
[29]Institute for Molecular Medicine Finland, FIMM, Helsinki 00290, Finland
[30]Department of Gastroenterology, Hepatology and Immunology, Children's Memorial Health Institute, Warsaw, Poland
[31]European Laboratory for Food Induced Disease, University of Naples Federico II, Naples, Italy
[32]Department of Gastroenterology and Hepatology, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands
[33]Department of Pathology, Children's Memorial Health Institute, Warsaw, Poland
[34]Department of Population Sciences, Beckman Research, Institute of the City of Hope, Duarte, CA, USA
[35]Department of Experimental Medicine, Faculty of Medicine University of Milano-Bicocca, Monza, Italy
[36]Department of Medical Genetics & Research Program for Molecular Medicine, Biomedicum Helsinki, University of Helsinki, Helsinki, Finland
[37]These authors contributed equally

Nature Genetics (2010) aceptado

Resumen

Llevamos a cabo un estudio de barrido genómico de segunda generación con 4.533 pacientes con enfermedad celíaca y 10.750 controles. Genotipamos 113 SNP seleccionados con $p_{GWAS}<10^{-4}$, y 18 SNPs de 14 loci conocidos en una muestra adicional de 4.918 casos y 5.684 controles.Variantes génicas de 13 nuevas regiones alcanzan el umbral de significación ($p_{combined}<5x10^{-8}$), conteniendo mayormente genes con función inmunológica (*BACH2, CCR4*, *CD80*, *CIITA*/*SOCS1*/*CLEC16A, ICOSLG*, *ZMIZ1*), jugando los genes *ETS1, RUNX3, THEMIS* y *TNFRSF14* un papel clave en en la selección tímica de células T. Además otras 13 regiones muestran evidencia sugerente de asociación. En un meta-análisis de rasgos cuantitativos de expresión de 1.469 muestras de sangre, 20 de 38 (52.6%) loci analizados tienen variantes de riesgo celíaco correlacionado ($p<0.0028$, FDR 5%) con una expresión génica en *cis*.

Abstract

We performed a second-generation genome wide association study of 4,533 celiac disease cases and 10,750 controls. We genotyped 113 selected SNPs with $p_{GWAS}<10^{-4}$, and 18 SNPs from 14 known loci, in a further 4,918 cases and 5,684 controls. Variants from 13 new regions reached genome wide significance ($p_{combined}<5x10^{-8}$), most contain immune function genes (*BACH2*, *CCR4*, *CD80*, *CIITA/SOCS1/CLEC16A*, *ICOSLG*, *ZMIZ1*) with *ETS1*, *RUNX3*, *THEMIS* and TNFRSF14 playing key roles in thymic T cell selection. A further 13 regions had suggestive association evidence. In an expression quantitative trait meta-analysis of 1,469 whole blood samples, 20 of 38 (52.6%) tested loci had celiac risk variants correlated ($p<0.0028$, FDR 5%) with cis gene expression.

Introduction

Celiac disease is a common heritable chronic inflammatory condition of the small intestine induced by dietary wheat, rye and barley, as well as other unidentified environmental factors, in susceptible individuals. Specific *HLA-DQA1* and *HLA-DQB1* risk alleles are necessary, but not sufficient, for disease development [1,2]. The well defined role of HLA-DQ heterodimers, encoded by these alleles, is to present cereal peptides to CD4+ T cells, activating an inflammatory immune response in the intestine. A single genome wide association study (GWAS) has been performed in celiac disease, which identified the *IL2/IL21* risk locus [1]. Subsequent studies probing the GWAS information in greater depth have identified a further 12 risk regions. Most of these regions contain a candidate gene functional in the immune system, although only in the case of *HLA-DQA1* and *HLA-DQB1* have the causal variants been established [3-5]. Many of the known celiac loci overlap with other immune-related diseases [6]. In order to identify additional risk variants, particularly of small effect size, we performed a second-generation GWAS using over six times as many samples as the previous GWAS and a denser genome-wide SNP set. We followed up promising findings in a large collection of independent samples.

The GWAS included five European celiac disease case and control sample collections including the previously reported celiac disease dataset [1]. We performed stringent data quality control (Online Methods), including calling genotypes using a custom algorithm on both large sample sets, and where possible cases and controls together (Online Methods). We tested 292,387 non-HLA SNPs from the Illumina Hap300 marker set for association in 4,533 celiac disease cases and 10,750 controls of European descent (Table 1). A further 231,362 additional non-HLA markers from the Illumina Hap550 marker set were tested for association in a subset of 3,796 celiac disease cases and 8,154 controls. All markers were from autosomes or the X chromosome. Genotype call rates were >99.9% in both datasets. The overdispersion factor of association test statistics, λ_{GC}=1.12, was similar to that observed in other GWAS of this sample size [7,8]. Findings were not substantially altered by imputation of missing genotypes for 737 celiac disease cases genotyped on the Hap300 BeadChip and corresponding controls (Table 1, collection 1). Here we present results for directly genotyped samples, as around half the additional Hap550 markers cannot be

accurately imputed from Hap300 data [9] (including the new *ETS1* locus finding in this study). Results for the top 1000 markers are available in Supplementary Data 1 *(CD adjunto)*, but because of concerns regarding identity detection of individuals [10], results for all markers are available only on request to the corresponding author.

For follow-up, we first inspected genotype clouds for the 417 non-HLA SNPs meeting $p_{GWAS}<10^{-4}$, being aware that top GWAS association signals may be enriched for genotyping artefact, and excluded 22 SNPs from further analysis using a low threshold for possible bias. We selected SNPs from 113 loci for replication. Markers that passed design and genotyping quality control included: a) 18 SNPs from all 14 previously identified celiac disease risk loci (including a tag SNP for the major celiac disease associated HLA-DQ2.5cis haplotype [1]); b) 13 SNPs from all 7 novel regions with $p_{GWAS} <5x10^{-7}$; c) 86 SNPs from 59 of 68 novel regions with $5x10^{-7}> p_{GWAS} >5x10^{-5}$ in stage 1; d) 14 SNPs from 14 of 30 novel regions with $5x10^{-5}> p_{GWAS} >10^{-4}$ in stage 1 (for this last category, we mostly chose regions with immune system genes). Two SNPs were selected per region for: regions with stronger association; regions with possible multiple independent associations; and/or containing genes of obvious biological interest. We successfully genotyped 131 SNPs in 7 independent follow-up cohorts comprising 4,918 celiac disease cases and 5,684 controls of European descent. Genotype call rates were >99.9% in each collection. Primary association analyses of the combined GWAS and follow-up data were performed with a two-sided 2x2x12 Cochran-Mantel-Haenszel test.

Table 1: Sample collections and genotyping platforms.

Collection	Country	Celiac disease cases			Controls		
		Sample size (pre-QC)[a]	Sample size (post-QC)[b]	Platform[c]	Sample size (pre-QC)[a]	Sample size (post-QC)[b]	Platform[c]
Stage 1: Genome wide association							
1[ef]	UK	778	737	Illumina Hap300v1-1	2,596[j]	2,596	Illumina Hap550-2v3
2[eg]	UK	1,922	1,849	Illumina 670-QuadCustom_v1	5,069[j]	4,936	Illumina 1.2M-DuoCustom_v1
3[e]	Finland	674	647	Illumina 670-QuadCustom_v1	1,839[j]	1,829	Illumina 610-Quad
4[eh]	Netherlands	876	803	Illumina 670-QuadCustom_v1	960	846	Illumina 670-QuadCustom_v1
5[e]	Italy	541	497	Illumina 670-QuadCustom_v1	580	543	Illumina 670-QuadCustom_v1
Analysis of Hap300 markers			**4,533**			**10,750**	
Analysis of additional Hap550 markers			**3,796**			**8,154**	
Stage 2: Follow-up							
6	USA	987	973	Illumina GoldenGate	615	555	Illumina GoldenGate
7	Hungary	979	965	Illumina GoldenGate	1,126	1,067	Illumina GoldenGate
8[ei]	Ireland	653	597	Illumina GoldenGate	1,499	1,456	Illumina GoldenGate
9	Poland	599	564	Illumina GoldenGate	745	716	Illumina GoldenGate
10	Spain	558	550	Illumina GoldenGate	465	433	Illumina GoldenGate
11[e]	Italy	1,056	1,010	Illumina GoldenGate	864	804	Illumina GoldenGate
12[e]	Finland	270	259	Illumina GoldenGate	653[j]	653	Illumina 610-Quad[d]
Subtotal			**4,918**			**5,684**	
Analysis of Hap300 markers, and follow-up (91 SNPs)			**9,451**			**16,434**	
Analysis of additional Hap550 markers, and follow-up (40 SNPs)			**8,714**			**13,838**	

[a]Sample numbers attempted for genotyping, before any quality control (QC) steps were applied.
[b]Sample numbers after all quality control (QC) steps (used in the association analysis).
[c]All platforms contain a common set of Hap300 markers; the Hap550, 610-Quad, 670-Quad and 1.2M contain a common set of Hap550 markers.
[d]Finnish stage 2 controls were individuals within the Finrisk collection for whom Illumina 610-Quad genotype data became available after the completion of stage 1.
[e]As an additional quality control step, we performed case-case and control-control comparisons for collection 1 versus 2, and collection 3 versus 12, for the 40 SNPs in **Table 2** and observed no markers with $P<0.01$. We did observe (as expected) differences for collection 5 versus 11, from Northern and Southern Italy, respectively.
[f]All 737 post-QC cases reported in a previous GWAS[1].
[g]690 of the post-QC cases and 1150 of the post-QC controls were included in a previous GWAS follow-up study[22].
[h]498 of the post-QC cases and 767 of the post-QC controls were included in a previous GWAS follow-up study[22].
[i]352 of the post-QC cases and 921 of the post-QC controls were included in a previous GWAS follow-up study[22].
[j]Some of these data were generated elsewhere, and some prior quality control steps (information not available) had been applied.

Online Methods

Subjects:

Written informed consent was obtained from all subjects, with Ethics Committee / Institutional Review Board approval. All individuals are of European ancestry. Affected celiac individuals were diagnosed according to standard clinical, serological and histopathological criteria including small intestinal biopsy. DNA samples were from blood, lymphoblastoid cell lines or saliva.

GWAS: UK(1) celiac cases were previously described [1], with removal of additional individuals showing genetic relatedness to UK(2) individuals, and were matched to 1,958 Birth Cohort population controls genotyped by the Type 1 Diabetes Genetics Consortium [7]. Details of Dutch and Italian cohorts were previously described [18,35]. UK(2) celiac cases were partly previously described [22], with additional individuals recruited through Celiac UK membership. UK(2) controls comprised 2,434 population controls from the 1,958 Birth Cohort and 2.502 National Service population controls. Finnish affected individuals (sporadic cases, or unrelateds from affected families across Finland) were partly previously described [36]. Finnish population controls comprised 904 samples from Finrisk (Corogene, excluding coronary heart disease) and 925 samples from Health 2000 (excluding metabolic syndrome and positive celiac disease serology).

Follow-up: USA comprised 525 celiac cases and 340 controls from the Mayo Clinic (Minnesota), and 448 celiac cases and 215 controls from the University of California Irvine [33]. Polish celiac cases were diagnosed in hospital clinics, and controls from donors at the Children's Memorial Health Institute (Warsaw), excluding celiac serology positive samples. Italian samples comprised 377 celiac cases and 94 controls from Rome, and 637 celiac cases and 711 celiac serology negative controls from Naples [37]. Irish celiac cases and controls were as described, with additional samples [38]. 259 Finnish celiac cases were recruited similarly to GWAS samples, and controls were an additional 653 population controls from the Finrisk study. 965 Hungarian celiac cases were collected from Budapest and Debrecen children clinic, and 1.067 controls representative of the Hungarian population were selected from an epidemiological study. Part of the Hungarian cohorts have been described earlier [36].

Spanish celiac cases were recruited in Madrid hospitals, controls were donors and hospital employees [39].

GWAS genotyping:

See Table 1 for genotyping platform details. UK(1) case and control genotyping was previously described [1,7]. Illumina 670-Quad (a custom chip designed for the WTCCC2 and comprising Hap550 and common CNV content) and 610-Quad genotyping was performed for this study at Barts and The London Genome Centre, Wellcome Trust Sanger Institute and University Medical Centre Groningen. Illumina 1.2M-Duo (a custom chip designed for the WTCCC2 and comprising Illumina 1M and common CNV content) genotyping was performed at the Wellcome Trust Sanger Institute for the WTCCC2. Bead intensity data was processed and normalized for each sample in BeadStudio, R and theta values exported and genotype calling performed using an algorithm as described [1,40].

To minimize calling biases arising from batch effects and the use of different genotyping platforms, we attempted to call cases and controls together from the same collections if possible, and maximized samples size to reduce problems when clustering rare variants. Genotypes were called separately in the following five pools from normalized R, theta data: pool A (n=2,199: 778 UK1 cases together with 1,421 1.958 birth cohort controls genotyped on the Illumina Hap550v1-1 platform as described in our previous study [1]. Genotype data from the 1.421 controls was then discarded and UK1 case data subsequently merged with UK1 controls for analysis. UK1 controls were called separately in pool B (n=2,596, 1,958 birth cohort individuals genotyped for the Type 1 Diabetes Genetics Consortium on the Illumina Hap550-2v3 platform) [7]. Separate calling of UK1 case and UK1 control individuals was required due to differences in genotype cluster intensity characteristics observed for UK1 case data (Hap300v1-1) and controls (Hap550-2v3). Pool C comprised UK3 cases and controls (n=6,963, 1,894 cases on Human 670-QuadCustom_v1 and 5,069 controls on 1.2M-DuoCustom_v1). Pool D comprised Dutch and Italian cases and controls (n=2,917 on 670-QuadCustom_v1). Pool E comprised Health2000 and Finrisk cohorts and Finnish cases (n=6,760, 674 cases on Human 670-QuadCustom_v1, 912 Finrisk controls, 927 Health2000 controls, 4,247 additional

Finrisk and H2000 cases and internal controls- Human 610-Quad). The 4,247 additional samples were then discarded.

Quality control steps were performed in the following order: Very low call rate samples and SNPs were first excluded. SNPs were excluded from all sample collections if any collection showed call rates were <95% or deviation from Hardy-Weinberg equilibrium ($p<0.0001$) in controls. Samples were excluded for call rate <98%, incompatible recorded gender and genotype inferred gender, ethnic outliers (identified by multi-dimensional scaling plots of samples merged with HapMap Phase II data), duplicates and first degree relatives. 22 of 417 SNPs showing apparent association ($p_{GWAS} <10^{-4}$) were excluded after visual inspection of R theta plots suggested possible bias.

The over-dispersion factor of association test statistics (genomic control inflation factor), λ_{GC}, was calculated using observed versus expected values for all SNPs in PLINK.

Follow-up genotyping:

See Table 1. Finnish controls [12] were genotyped on the 610-Quad BeadChip at the Wellcome Trust Sanger Institute. All other samples were genotyped using the Illumina GoldenGate assay on the Veracode/BeadXpress platform at Barts and The London Genome Centre; King's College London; and University Medical Centre Groningen. Genotyping calling was performed in BeadStudio for combined cases and controls in each separate collection, with the exception of the Finnish collection, and whole genome amplified samples (89 Irish cases and 106 Spanish controls). Quality control steps were performed as for the GWAS. 131 of 144 SNPs passed quality control and visual inspection of genotype clouds.

SNP association analysis:

Analyses were performed using PLINK v1.07 [41]. The Cochran-Mantel-Haenzel test was used for most analyses. Logistic regression analyses were used to define the independence of association signals within the same linkage disequilibrium block, with group membership included as a factorized covariate.

Genotype imputation was performed for samples genotyped on the Illumina Hap300 using BEAGLE and the CEU, TSI, MEX and GIH reference samples from HapMap3. Association analysis was performed using logistic regression on posterior genotype probabilities, with group membership included as a factorized covariate.

The fraction of additive variance was calculated using a liability threshold model [42] and assuming a population prevalence of 1%. Effect sizes and control allele frequencies were estimated only from the combined replication panel. Genetic variance was calculated assuming 50% heritability.

GRAIL analysis:

We performed GRAIL analysis (http://www.broadinstitute.org/mpg/grail/grail.php) using HG18 and Dec2006 PubMed datasets, default settings for SNP rs number submission, and the 27 genome-wide significant celiac disease risk loci (most associated SNP) as seeds. As a query we used either associated SNPs, or 101 x 50 randomly chosen Hap550 SNP datasets (5,050 SNPs, of which 5,033 mapped to the GRAIL database).

Identification of Transcriptional Components:

We made the observation that the power of eQTL studies in humans is limited by substantial observed inter-individual variation in expression measurements due to non-genetic factors, and therefore developed a method, Transcriptional Components, to remove a large component of this variation (manuscript in preparation). Expression data from 42,349 heterogeneous human samples hybridized to Affymetrix HG-U133A (GEO accession number: GPL96) or HG-U133 Plus 2.0 (GEO accession number: GPL570) Genechips were downloaded [43] (Fig. 1, step 1). Samples missing data for >150 probes were excluded, and only probes available on both platforms were analysed, resulting in expression data for 22,106 probes and 41,408 samples. We performed quantile normalization using the median rank distribution [44] and log2 transformed the data - this ensured an identical distribution of expression signals for every sample, discarding previous normalization and transformation steps.

Initial quality control (QC) was performed by applying principal component analysis (PCA) on the sample correlation matrix (pair-wise Pearson correlation coefficients between all samples). The first principal component (PC), explaining ~80-90% of the total variance [45,46], describes probe-specific variance. 6,375 samples with correlation $R < 0.75$ of the sample array with this PC were considered outliers of lesser quality and excluded from analysis. We excluded entire GEO datasets where >25% of the samples were outliers (probably expression ratios versus a reference, not absolute data). The final dataset comprised 33,109 samples (17,568 GPL96 and 15,541 GPL570 samples), and repeated the normalization and transformation on the originally deposited expression values of these post-quality control samples.

We next applied PCA on the pairwise 22,106 x 22,106 probe Pearson correlation coefficient matrix assayed on the 33,109 sample dataset (we developed a speed-optimized C++ software tool, MATool, available upon request), attempting to simplify the structure of the data. Here, PCA represents a transformation of a set of correlated probes into sets of uncorrelated linear additions of probe expression signals (eigenvectors) that we name Transcriptional Components (TCs). Each TC is a weighted sum of probe expression signals and eigenvector probe coefficients, and TC-scores can be calculated for each observed expression array sample. A TC-score can be seen as the activity of the TC in a sample.

Subjects for expression - genotype correlation:

We obtained peripheral blood DNA and RNA (PAXgene) from Dutch and UK individuals who were disease cases or controls for GWAS studies (Supplementary Table 1). All samples had been genotyped for a common SNP set on Illumina platforms. Analysis was confined to 294,767 SNPs that had a MAF $\geq$ 5%, call-rate $\geq$ 95% and exact HWE $P>0.001$. RNA from the samples was either hybridized to Illumina HumanRef-8 v2 arrays (229 samples, Ref-8v2) or Illumina HumanHT-12 arrays (1,240 samples, HT-12), and raw probe intensity extracted using BeadStudio. The Ref-8v2 samples were jointly quantile normalized and log2 transformed, and similarly for the HT-12 samples. Subsequent analyses were also conducted separately for these datasets, up to the eventual eQTL mapping, that uses a meta-analysis framework, combining eQTL results from both arrays. HT-12 and Ref-8v2 arrays are

different, but share many probes with identical probe sequences. Illumina sometimes use different probe identifiers for the same probe sequences - in meta-analysis and Table 3, the label HT-12 was used if both HT-12 and Ref-8v2 had the same sequence.

Supplementary Table 1 Subjects used in eQTL mapping analysis.Numbers are shown after quality control steps.

Case/Control	Population	Number of individuals	Genotype platform	Expression platform
Controls	Dutch	324	Hap370	HT-12
Amyotrophic Lateral Sclerosis	Dutch	414	Hap370	HT-12
Ulcerative Colitis	Dutch	49	610-Quad	HT-12
Chronic Obstructive Pulmonary Disease	Dutch	453	610-Quad	HT-12
Celiac disease	UK	111	Hap300	Ref-8 v2
Amyotrophic Lateral Sclerosis	Dutch	59	Hap300	Ref-8 v2
Controls	Dutch	59	Hap300	Ref-8 v2

Re-mapping of probes:

If probes mapped incorrectly, or cross-hybridized to multiple genomic loci, it might be that an eQTL would be detected that would be deemed a trans-eQTLs. To prevent this, we used a mapping approach we developed for RNA-seq data [47]. We used Ensembl52 to obtain, for each annotated gene, the transcript with the largest number of exons and included this main spliced transcript in our reference set. Second, we added one sequence per intron, extending intron boundaries 40 bp on each side to allow mapping of the 50 bp probe sequences, overlapping exon-intron junctions. Last, a version of the reference DNA genome with masked annotated transcripts was included. Probe sequences were mapped versus the composite reference using novoalign v2.05.12 permitting alignment against all possible sequences originating from the same transcript (parameters -t 150 -v 20 20 200 [>]([^_]*)_). For each probe we determined whether it mapped uniquely to one particular genomic locus, or, if multiple hits were present (e.g. across exons) that these resided < 250kb from the probe. Probes that did not map, or mapped to multiple different locations were removed.

Affymetrix transcriptional components applied to Illumina expression data:

TC can be applied to new datasets to calculate the TC-scores for each new individual sample. In order to calculate the TC-scores for the Illumina samples used for the cis-eQTL mapping, only Illumina probes that could be mapped to any of our 22,106 Affymetrix probes were used (www.switchtoi.com/probemapping.ilmn). The TC-score of sample i for the jth TC is defined as: $TCscore_{ij} = \sum_{t=1}^{t=n} a_{ti} \times v_{tj}$, where *v*tj is defined as the tth probe coefficient for the jth TC; ati is the expression measurement for the tth mapped probe for sample i.

Removal of transcriptional component effects from Illumina expression data:

We correlated the TC-scores for each peripheral blood sample with probe expression levels (Fig. 1), and then used the residual gene expression data for subsequent cis-eQTL mapping.

***cis*-eQTL zapping:**

We used the residual gene expression data (Fig. 1) in a meta-analysis framework, as described [48,49]. In brief, analyses were confined to those probe-SNP pairs for which the distance from probe transcript midpoint to SNP genomic location was less than 500 kb. To prevent spurious associations due to outliers, a non-parametric Spearman's rank correlation analysis was performed. When a particular probe-SNP pair was present in both the HT12 and H8v2 datasets, an overall, joint p-value was calculated using a weighted Z-method (square root of the dataset sample number). To correct for multiple testing we controlled the false discovery rate (FDR). The distribution of observed P values was used to calculate the FDR, by permuting expression phenotypes relative to genotypes 1000 times within the HT12 and H8v2 dataset. Finally, we removed probes containing a known SNP which may give artefactual results, using 1000Genomes CEU SNP data (April 2009 release).

Results

Celiac disease risk variants

The *HLA* locus and all 13 other previously reported celiac disease risk loci showed evidence for association at a genome wide significance threshold ($p_{combined}<5x10^{-8}$) (Table 2). We note that some loci were previously reported using less stringent criteria (i.e. the P<5x10-7 recommended by the 2007 WTCCC study [11]), but that in the current, much larger sample set, all known loci meet recently proposed $p<5x10^{-8}$ thresholds [12,13].

We identified 13 novel risk regions with genome-wide significant evidence (Pcombined <5x10-8) of association, including regions containing the *BACH2, CCR4, CD80, CIITA/SOCS1/CLEC16A, ETS1, ICOSLG, RUNX3, THEMIS, TNFRSF14*, and *ZMIZ1* genes which are of obvious immunological function (Table 2). A further 13 regions met 'suggestive' criteria for association (either $10^{-6}< P_{combined} <5x10^{-8}$ and/or PGWAS<10-4 and Pfollowup<0.01). These regions also contain multiple genes of obvious immunological function, including *CD247, FASLG/TNFSF18/TNFSF4, IRF4, TLR7/TLR8, TNFRSF9* and *YDJC*. Six of the 39 non-HLA regions show evidence for the presence of multiple independently associated variants in a conditional logistic regression analysis (Supplementary Data 2 *(CD adjunto)*).

We tested the 40 SNPs with the strongest association (Table 2) from each of the known genome-wide significant, new genome-wide significant, and new suggestive loci for evidence of heterogeneity across the 12 collections studied. Only the HLA region was significant (Breslow-Day test p<0.05 / 40 tests, rs2187668 $p=4.8x10^{-8}$) which is consistent with the well described North-South gradient in HLA allele frequency in European populations, and more specifically for HLA-DQ in celiac disease [14].

We observed no evidence for interaction between each of the 26 genome-wide significant non-HLA loci, which is consistent with what has been reported for complex diseases so far. However, we did observe weak evidence for lower effect sizes at non-HLA loci in high risk HLA-DQ2.5*cis* homozygotes, similar to what has been previously observed in type 1 diabetes [7].

To obtain more insight into the functional relatedness of the celiac loci, we applied GRAIL, a statistical tool that utilizes text mining of PubMed abstracts to annotate candidate genes from loci associated with common disease risk [15,16]. We first performed a 'leave-one-out' analysis of the 27 genome-wide significant celiac disease loci (including HLA-DQ) and obtained GRAIL scores of $p_{text}<0.01$ for 12 loci (44% sensitivity, Table 2). Factors that limit the sensitivity of GRAIL include biological pathways being both known (a 2006 dataset is used to avoid GWAS era studies), and published in the literature. We then applied GRAIL analysis, using the 27 known regions as a seed, to all 49 regions (49 SNPs) with 10^{-3} >Pcombined >$5x10^{-8}$ and obtained GRAIL p_{text} <0.01 for 9 regions (18.4%). As a control, only 5.5% (279 of 5,033) of randomly selected Hap550 SNPs reached this threshold. In addition to the five 'suggestive' loci shown in Table 2, GRAIL annotated four further interesting gene regions of lower significance in the combined association results: rs944141/*PDCD1LG2* ($p_{combined}=4.4x10^{-6}$), rs976881/*TNFRSF8* ($p_{combined}=2.1x10^{-4}$), rs4682103/*CD200*/*BTLA*($p_{combined}=6.8x10^{-6}$) and rs4919611/*NFKB2* ($p_{combined}=6.1x10^{-5}$). There appeared to be further enrichment for genes of immunological interest which are not GRAIL annotated in the $10^{-3}>p_{combined}>5x10^{-8}$ significance window, including rs3828599/*TNIP1* ($p_{combined}=1.55x10^{-4}$), rs8027604/*PTPN9* ($p_{combined}=1.4x10^{-6}$), rs944141/*CD274* ($p_{combined}=4.4x10^{-6}$). Some of these findings, for which neither genome-wide significant nor suggestive association is achieved, are likely to comprise part of a longer tail of disease predisposing common variants, of weaker effect sizes. Definitive assessment of these biologically plausible regions would require genotyping and association studies using much larger sample collections than the present study.

We previously showed considerable overlap between celiac disease and type 1 diabetes risk loci [17], as well as celiac disease and rheumatoid arthritis [18], and more generally, there is now substantial evidence for shared risk loci between the common chronic immune mediated diseases [6]. To update these observations, we searched 'A Catalog of Published Genome Wide Association Studies' (18 Nov 2009) [19] and the HuGE database [20]. We found some evidence (requiring a published association report of $p<1x10^{-5}$) of shared loci with at least one other inflammatory or immune mediated disease for 18 of the current 27 genome-wide significant celiac risk regions. We defined shared regions as the broad LD block, however different SNPs

are often reported in different diseases, and at only three of the 18 shared regions are associations across all diseases with the same SNP or a proxy SNP in $r^2>0.8$ in HapMap CEU. Currently 9 regions appear celiac disease specific and may reflect distinctive disease biology including the regions containing rs296547 and rs9792269, and the regions around *CCR4, CD80, ITGA4, LPP, PLEK, RUNX3* and *THEMIS*. In fact, sharing is probably greater, due to both stochastic variation in results from sample size limitations, and regions with a genuinely stronger effect size in one disease and weaker effect size in another.

Genetic variation in *ETS1* has recently been reported to be associated with systemic lupus erythematosus (SLE) in the Chinese population, although is not associated with SLE in European populations [21]. The most strongly associated celiac (European population) SNP rs11221332 and the most strongly associated SLE (Chinese population) SNP rs6590330 map 70kb apart. Inspection of the HapMap phase II data shows broadly similar linkage disequilibrium patterns between Chinese (CHB) and European (CEU) populations in this region, with the two associated SNPs in non-adjacent linkage disequilibrium blocks. Thus distinct common variants within the same gene are predisposing to different autoimmune diseases across different ethnic groups.

Function of celiac risk variants

Celiac risk variants in the HLA alter protein structure and function [4]. However we identified only four non-synonymous SNPs with evidence for celiac disease association ($p_{GWAS}<10^{-4}$) from the other 26 genome-wide significant associated regions (rs3748816/*MMEL1*, rs3816281/*PLEK*, rs196432/*RUNX3*, rs3184504/*SH2B3*). Although comprehensive regional resequencing is required to test the possibility that coding variants contribute to the observed association signals, more subtle effects of genetic variation on gene expression are the more likely major functional mechanism for complex disease genes. With this in mind, we performed a meta-analysis of new and published genome-wide expression quantitative trait loci (eQTL) datasets comprising 1,469 human whole blood (PAXgene) samples reflecting primary leucocyte gene expression. We applied a new method, Transcriptional Components, to remove a substantial proportion of inter-individual non-genetic expression variation

and performed eQTL meta-analysis on the residual expression variation (Online Methods).

We assessed 38 of the 39 genome-wide significant and suggestive celiac disease associated non-HLA loci (Table 2) for cis expression - genotype correlations. We tested the SNP with the strongest association from each region. However for five regions the most associated SNP was not genotyped in the eQTL samples (Hap300 data), instead for four of these we tested a proxy SNP (r^2>0.5 in HapMap CEU). In addition, for six loci showing evidence of multiple independent associations in conditional regression analyses, we tested a second SNP showing independent celiac disease association for eQTL analysis. In total we assessed 44 independent non-HLA SNP associations in peripheral whole blood samples genotyped on the Illumina Hap300 BeadChip and either Illumina Ref8 or HT12 expression arrays, correlating each SNP with data from gene-probes mapping within a 1Mb window.

We identified significant (Spearman p<0.0028, corresponding to 5% false discovery rate) eQTLs at 20 of 38 (52.6%) non-HLA celiac loci tested (Table 3, Supplementary Figures 2 & 3 *(CD adjunto)*). Some loci had evidence of eQTLs with multiple probes, genes or SNPs (Table 3). We assessed whether the number of SNPs with *cis*-eQTL effects out of the 44 SNPs that we tested, was significantly higher than expected. We observed that eQTL SNPs on average have a substantially higher MAF than non-eQTL SNPs in the 294,767 SNPs tested. In order to correct for this we selected 44 random SNPs that had an equal MAF distribution, and determined for how many of these MAF-matched SNPs eQTLs were observed. We observed a significantly higher number of eQTL SNPs ($p=9.3 \times 10^{-5}$, 106 permutations) amongst the celiac associated SNPs than expected by chance (22 observed eQTL SNPs, vs. 7.8 expected eQTL SNPs). Therefore the celiac disease associated regions are greatly enriched for eQTLs. These data suggest some risk variants may influence celiac disease susceptibility through a mechanism of altered gene expression. Biologically plausible candidate genes with a significant eQTL, where the peak eQTL signal and peak case/control association signal are similar (Supplementary Figure 3 *(CD adjunto)*), include *CD247, IL18RAP* (previously reported [22]), *PARK7, PLEK, TAGAP* and *ZMIZ1*.

We also assessed co-expression of genes mapping within 500kb of SNPs showing strongest case/control association from the 40 genome-wide significant and suggestive celiac disease loci in an analysis of the 33,109 human Affymetrix Gene Expression Omnibus dataset. This analysis loses power to detect tissue specific correlations by use of numerous tissue types, but greatly gains power by the large sample size. We detected several distinct co-expression clusters (Pearson correlation coefficient between genes >0.5), including four clusters of immune-related genes which contain at least one gene from 37 of the 40 genome-wide significant and suggestive loci (Fig. 1). These data further demonstrate that genes from celiac disease risk loci map to multiple distinct immunological pathways involved in disease pathogenesis.

Tabla 2

Tabla 3

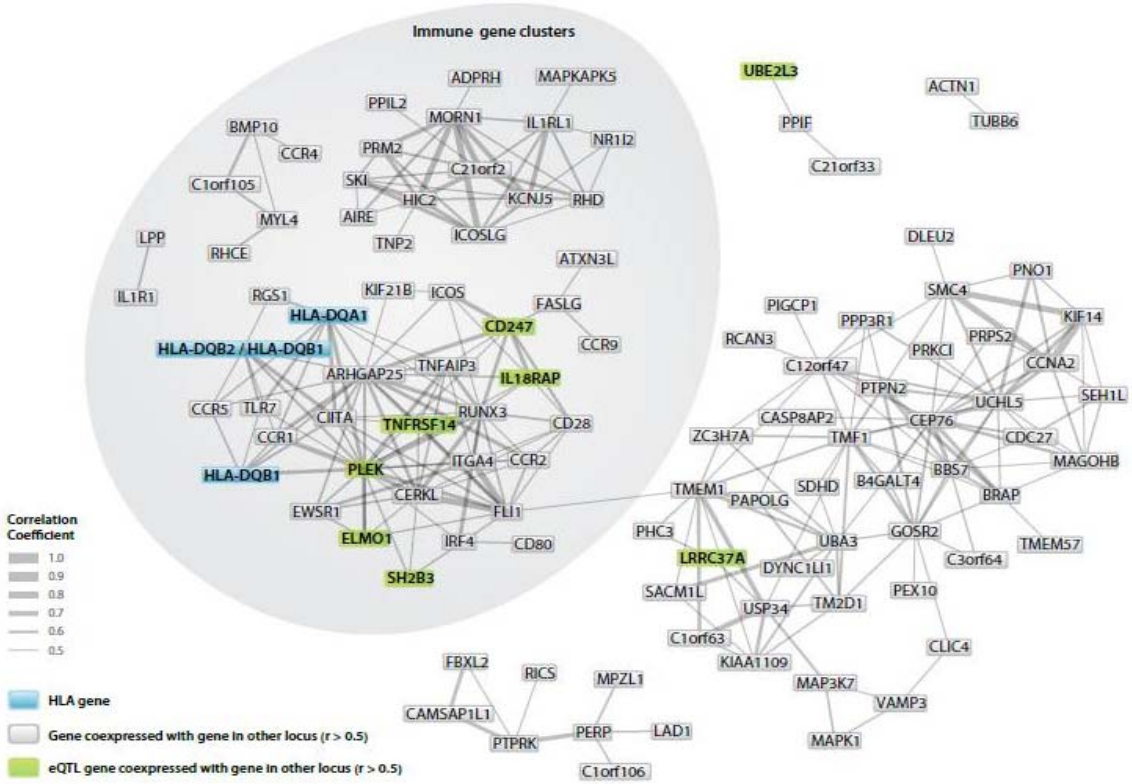

Figure 1: Co-expression analysis of genes mapping to 40 genome-wide significant and suggestive celiac disease regions in 33,109 heterogenous human samples from the Gene Expression Omnibus. Genes mapping within a 1Mb window of associated SNPs (Table 2) were tested for interaction with genes from other loci. Interactions with Pearson correlation >0.5 shown ($P<10^{-100}$). Only the genes known to contain causal mutations (*HLA-DQA1*, *HLA-DQB1*) were analysed from the HLA region, "HLA-DQB2 / HLA-DQB1" is a single expression probe mapping to both genes. No probe for *THEMIS* was present on the earlier version of the U133 array, however in a subset analysis of U133 Plus data, *THEMIS* is co-expressed in the major immune gene cluster.

Discussion

We previously reported that most celiac genetic risk variants mapped near genes that are functional in the immune system [22], and this remains true for the 13 new genome-wide significant, and 13 new suggestive, risk variants from the current study. We can now refine these observations and highlight specific immunological pathways relevant to celiac disease pathogenesis:

1) T cell development in the thymus. The rs802734 LD block contains the recently identified gene *THEMIS* 'THymus-Expressed Molecule Involved in Selection'. THEMIS plays a key regulatory role in both positive and negative T-cell selection during late thymocyte development [23]. Furthermore, the rs10903122 LD block contains *RUNX3*, a master regulator of CD8+ T lymphocyte development in the thymus[193, 194]. *TNFRSF14* (*LIGHTR*, rs3748816 LD block) has widespread peripheral leucocyte functions as well as a critical role in promoting thymocyte apoptosis [26]. The *ETS1* transcription factor (rs11221332 LD block) is also active in peripheral leucocytes, however it is also a key player in thymic CD8+ lineage differentiation, acting in part by promoting *RUNX3* expression [27].

The importance of the thymus in autoimmune disease pathogenesis has been previously emphasised by the established role of thymectomy in the treatment of myasthenia gravis. In type 1 diabetes, it was shown that disease associated genetic variation in the insulin gene *INS* causes altered thymic insulin expression and subsequent T cell tolerance for insulin as a self-protein [28]. However, the importance of thymic T cell regulation has not been previously recognised in the aetiology of celiac disease. It is conceivable that the associated variants may alter biological processes prior to thymic MHC-ligand interactions. Alternatively it is now clear that exogenous antigen presentation and selection occurs in the thymus via migratory dendritic cells - this has been demonstrated for skin, and has been hypothesised for food antigens [29,30]. These findings suggest research into immuno-/pharmaco-logical modifiers of T cell tolerance more generally in autoimmune diseases.

2) Innate immune detection of viral RNA. Although the association signal at rs5979785 ($p_{combined}=6.36x10^{-8}$) in the *TLR7/TLR8* region is just outside our genome wide significance threshold, we observe a strong effect of rs5979785 on *TLR8*

expression in whole blood. TLR8 recognises viral single stranded RNA. Taken together with the recent observation of rare loss of function mutations in the enteroviral response gene *IFIH1* in type 1 diabetes [31], these findings suggest viral infection (and the nature of the host response to infection) as a putative environmental trigger common to these autoimmune diseases.

3) T and B cell co-stimulation (or co-inhibition). This class of molecules controls the strength and nature of the response to T or B (immunoglobulin) cell receptor activation by antigens. We observe multiple regions with genes (*CTLA4/ICOS/CD28, TNFRSF14, CD80, ICOSLG, TNFRSF9, TNFSF4*) from this class of ligand-receptor pairs suggesting fine control of the adaptive immune response might be altered in at-risk individuals.

4) Cytokines, chemokines and their receptors. Our previous report discussed the function of the 2q11-12 interleukin receptor cluster (*IL18RAP*, etc), the 3p21 chemokine receptor cluster (*CCR5*, etc.) and the loci containing *IL2/IL21* and *IL12A* [22]. We now report additional loci containing *TNFSF18* and *CCR4*.

We estimate that the current celiac disease variants, including the major celiac disease associated *HLA* variant, HLA-DQ2.5cis, less common celiac disease associated haplotypes in the *HLA* (HLA-DQ8; HLA-DQ2.5trans; HLADQ2.2), and the additional 26 definitively implicated loci explain about 20% of total celiac disease variance, which would represent 40% of genetic variance, assuming a heritability of 0.5. A long tail of low effect size common variants, along with highly penetrant rare variants (both at the established loci and elsewhere in the genome), may substantially contribute to the remaining heritability.

We observed different haplotypes within the *ETS1* region associated with coeliac disease in Europeans, and SLE in the Chinese population. We further note for some autoimmune diseases studied in European origin populations, that although the same LD block has been associated, the association is with a different haplotype. In some cases, the same variants are associated, but the direction of association is opposite (e.g. rs917997/*IL18RAP* in celiac disease versus type 1 diabetes). We believe further exploration of these signals may reveal critical differences in the nature of the immune system perturbation between these diseases.

Previous investigators have observed that only a small proportion of GWAS associations are coding variants, and have suggested that these may instead influence regulation of gene expression. Here, we show that over half the celiac disease associated variants are correlated with expression changes in nearby genes. This mechanism is likely to explain the function of some risk variants for other common, complex diseases. Further research, however, is needed to definitively determine at each locus both the celiac disease causal variants and their functional mechanisms.

References

[1] van Heel, D.A. et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* **39**, 827-9 (2007).

[2] van Heel, D.A. & West, J. Recent advances in coeliac disease. *Gut* **55**, 1037-46 (2006).

[3] Sollid, L.M. et al. Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. *J Exp Med* **169**, 345-50 (1989).

[4] Kim, C.Y., Quarsten, H., Bergseng, E., Khosla, C. & Sollid, L.M. Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. *Proc Natl Acad Sci U S A* **101**, 4175-9 (2004).

[5] Henderson, K.N. et al. A Structural and Immunological Basis for the Role of Human Leukocyte Antigen DQ8 in Celiac Disease. *Immunity* **27**, 23-34 (2007).

[6] Zhernakova, A., van Diemen, C.C. & Wijmenga, C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat Rev Genet* **10**, 43-55 (2009).

[7] Barrett, J.C. et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* (2009).

[8] Barrett, J.C. et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* **40**, 955-62 (2008).

[9] Anderson, C.A. et al. Evaluating the effects of imputation on the power, coverage, and cost efficiency of genome-wide SNP platforms. *Am J Hum Genet* **83**, 112-9 (2008).

[10] Jacobs, K.B. et al. A new statistic and its power to infer membership in a genome-wide association study using genotype frequencies. *Nat Genet* (2009).

[11] Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-78 (2007).

[12] Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M.J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* **32**, 381-5 (2008).

[13] Dudbridge, F. & Gusnanto, A. Estimation of significance thresholds for genome wide association scans. *Genet Epidemiol* **32**, 227-34 (2008).

[14] Karell, K. et al. HLA types in celiac disease patients not carrying the DQA1*05-DQB1*02 (DQ2) heterodimer: results from the European Genetics Cluster on Celiac Disease. *Hum Immunol* **64**, 469-77 (2003).

[15] Raychaudhuri, S. et al. Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk. *Nat Genet* (2009).

[16] Raychaudhuri, S. et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet* **5**, e1000534 (2009).

[17] Smyth, D.J. et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. *N Engl J Med* **359**, 2767-77 (2008).

[18] Coenen, M.J. et al. Common and different genetic background for rheumatoid arthritis and coeliac disease. *Hum Mol Genet* (2009).

[19] Hindorff, L.A. et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362-7 (2009).

[20] Yu, W., Clyne, M., Khoury, M.J. & Gwinn, M. Phenopedia and Genopedia: Disease-centered and Gene-centered Views of the Evolving Knowledge of Human Genetic Associations. *Bioinformatics* (2009).

[21] Han, J.W. et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat Genet* **41**, 1234-7 (2009).

[22] Hunt, K.A. et al. Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet* **40**, 395-402 (2008).

[23] Allen, P.M. Themis imposes new law and order on positive selection. *Nat Immunol* **10**, 805-6 (2009).

[24] Sato, T. et al. Dual functions of Runx proteins for reactivating CD8 and silencing CD4 at the commitment process into CD8 thymocytes. *Immunity* **22**, 317-28 (2005).
[25] Woolf, E. et al. Runx3 and Runx1 are required for CD8 T cell development during thymopoiesis. *Proc Natl Acad Sci U S A* **100**, 7731-6 (2003).
[26] Wang, J. & Fu, Y.X. LIGHT (a cellular ligand for herpes virus entry mediator and lymphotoxin receptor)-mediated thymocyte deletion is dependent on the interaction between TCR and MHC/self-peptide. *J Immunol* **170**, 3986-93 (2003).
[27] Zamisch, M. et al. The transcription factor Ets1 is important for CD4 repression and Runx3 up-regulation during CD8 T cell differentiation in the thymus. *J Exp Med* (2009).
[28] Vafiadis, P. et al. Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. *Nat Genet* **15**, 289-92 (1997).
[29] Bonasio, R. et al. Clonal deletion of thymocytes by circulating dendritic cells homing to the thymus. *Nat Immunol* **7**, 1092-100 (2006).
[30] Klein, L., Hinterberger, M., Wirnsberger, G. & Kyewski, B. Antigen presentation in the thymus for positive selection and central tolerance induction. *Nat Rev Immunol* **9**, 833-844 (2009).
[31] Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J.A. Rare Variants of IFIH1, a Gene Implicated in Antiviral Responses, Protect Against Type 1 Diabetes. *Science* (2009).
[32] Trynka, G. et al. Coeliac disease-associated risk variants in TNFAIP3 and REL implicate altered NF-kappaB signalling. *Gut* **58**, 1078-83 (2009).
[33] Garner, C.P. et al. Replication of celiac disease UK genome-wide association study results in a US population. *Hum Mol Genet* **18**, 4219-25 (2009).
[34] Plenge, R.M. et al. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* **39**, 1477-82 (2007).
[35] Romanos, J. et al. Six new coeliac disease loci replicated in an Italian population confirm association with coeliac disease. *J Med Genet* **46**, 60-3 (2009).
[36] Koskinen, L.L. et al. Fine mapping of the CELIAC2 locus on chromosome 5q31-q33 in the Finnish and Hungarian populations. *Tissue Antigens* **74**, 408-16 (2009).
[37] Megiorni, F. et al. HLA-DQ and susceptibility to celiac disease: evidence for gender differences and parent-of-origin effects. *Am J Gastroenterol* **103**, 997-1003 (2008).

[38] Hunt, K.A. et al. Large scale replication of a genome-wide association study in celiac disease. *American Society of Human Genetics Meeting*, platform talk 26 (2007).
[39] Dema, B. et al. Association of IL18RAP and CCR3 with coeliac disease in the Spanish population. *J Med Genet* **46**, 617-9 (2009).
[40] Franke, L. et al. Detection, imputation, and association analysis of small deletions and null alleles on oligonucleotide arrays. *Am J Hum Genet* **82**, 1316-33 (2008).
[41] Purcell, S. et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* **81**, 559-75 (2007).
[42] Risch, N.J. Searching for genetic determinants in the new millennium. *Nature* **405**, 847-56 (2000).
[43] Edgar, R., Domrachev, M. & Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207-10 (2002).
[44] Bolstad, B.M., Irizarry, R.A., Astrand, M. & Speed, T.P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-93 (2003).
[45] Sherlock, G. Analysis of large-scale gene expression data. *Brief Bioinform* **2**, 350-62 (2001).
[46] Alter, O., Brown, P.O. & Botstein, D. Singular value decomposition for genome-wide expression data processing and modeling. *Proc Natl Acad Sci U S A* **97**, 10101-6 (2000).
[47] Heap, G.A. et al. Genome-wide analysis of allelic expression imbalance in human primary cells by high throughput transcriptome resequencing. *Hum Mol Genet* (2009).
[48] Heap, G.A. et al. Complex nature of SNP genotype effects on gene expression in primary human leucocytes. *BMC Med Genomics* **2**, 1 (2009).
[49] Franke, L. & Jansen, R.C. eQTL analysis in humans. *Methods Mol Biol* **573**, 311-28 (2009).

ACKNOWLEDGMENTS

We thank Celiac UK for assistance with direct recruitment of celiac disease individuals, and UK clinicians (L.C. Dinesen, G.K.T. Holmes, P.D. Howdle, J.R.F. Walters, D.S. Sanders, J. Swift, R. Crimmins, P. Kumar, D. Jewell, S.P.L. Travis, K.

Moriarty) who recruited celiac disease blood samples described in our previous studies[126, 132]. We thank the genotyping facility of the UMCG (J. Smolonska , P. van der Vlies) for generating part of the GWAS and replication data and the gene expression data; R. Booij and M. Weenstra are thanked for preparation of Italian samples. H. Ahola, A. Heimonen, L. Koskinen, E. Einarsdottir and K. Löytynoja are thanked for their work on Finnish sample collection, preparation and data handling, and E. Szathmári, J.B.Kovács, M. Lörincz and A. Nagy for their work with the Hungarian families. Health2000 organization, Finrisk consortium, K. Mustalahti, M. Perola, K. Kristiansson and J. Koskinen are thanked for providing the Finnish control genotypes. We thank D.G. Clayton and N. Walker for providing T1DGC data in the required format. We thank the Irish Transfusion Service and Trinity College Dublin Biobank for control samples; V. Trimble, E. Close, G. Lawlor, A. Ryan, M. Abuzakouk, D. Kelleher, C. O'Morain, G. Horgan, for celiac sample collection and preparation We acknowledge DNA provided by Mayo Clinic Rochester, and Prof. M. Bonamico and Prof. M. Barbato (Department of Paediatrics, Sapienza University of Rome, Italy) for recruiting individuals. We thank Polish clinicians for recruitment of celiac disease individuals (Z. Domagala, A. Szaflarska-Poplawska, B. Oralewska, W. Cichy, B. Korczowski, K. Fryderek, E. Hapyn, K. Karczewska, A. Zalewska, I. Sakowska-Maliszewska, R. Mozrzymas, A. Zabka, M. Kolasa, B. Iwanczak). We thank M. Szperl for isolating DNA from blood samples provided by Children's Memorial Health Institute (Warsaw, Poland).

Dutch and UK genotyping for the second celiac disease GWAS was funded by the Wellcome Trust (084743 to D.A.vH.). Italian genotyping for the second celiac disease GWAS was funded by the Coeliac Disease Consortium, an Innovative Cluster approved by the Netherlands Genomics Initiative and partially funded by the Dutch Government (BSIK03009 to C.W.) and by the Netherlands Organisation for Scientific Research (NWO, VICI grant 918.66.620 to C.W.). E.G. is funded by the Italian Ministry of Healthy (grant RC2009). L.H.v.d.B. acknowledges funding from the Prinses Beatrix Fonds, the Adessium foundation, and the Amyotrophic Lateral Sclerosis Association. L.F. received a Horizon Breakthrough grant from the Netherlands Genomics Initiative (93519031) and a VENI grant from NWO (ZonMW grant 916.10.135). P.C.D. is a MRC Clinical Training Fellow (G0700545). G.T.

received a Ter Meulen Fund grant from the Royal Netherlands Academy of Arts and Sciences (KNAW). The gene expression study was funded in part by COPACETIC (EU grant 201379). This study makes use of data generated by the Wellcome Trust Case-Control Consortium 2 (WTCCC2). A full list of the WTCCC2 investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the WTCCC2 project was provided by the Wellcome Trust under award 085475. This research utilizes resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. We acknowledge the use of BRC Core Facilities provided by the financial support from the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London and King's College Hospital NHS Foundation Trust. We acknowledge funding from NIH: DK050678 and DK081645 (to S.L.N.), NS058980 (to R.A.O.); DK57892 and DK071003 (to J.A.M.). Collection of Finnish and Hungarian patients was funded by EU Commission (MEXT-CT-2005-025270), the Academy of Finland, Hungarian Scientific Research Fund (contract OTKA 61868), the University of Helsinki Funds, the Competitive Research Funding of the Tampere University Hospital, the Foundation of Pediatric Research, Sigrid Juselius Foundation, and the Hungarian Academy of Sciences (2006TKI247 to RA). Funding for the Polish samples collection and genotyping was provided by UMC Cooperation Project (6/06/2006/NDON). R.McM is funded by Science Foundation Ireland. C. Núñez has a FIS contract (CP08/0213). The Dublin Centre for Clinical Research contributed to patient sample collection and is funded by the Irish Health Research Board

Finally we thank all individuals with celiac disease and control individuals for participating in this study.

DISCUSIÓN GENERAL

La enfermedad celíaca (EC) es una de las enfermedades con componente autoinmune más frecuente, alcanzando una prevalencia del 1% en población caucásica. Se caracteriza por tener una etiología multifactorial, en la que tanto factores genéticos como ambientales influyen en la predisposición a la enfermedad. En individuos genéticamente susceptibles, la ingesta de alimentos ricos en gluten o proteínas relacionadas (prolaminas), desencadena el desarrollo de un proceso inflamatorio crónico en la mucosa intestinal, que en la gran mayoría de los casos conduce a la atrofia de las vellosidades intestinales.

Aunque el principal componente ambiental que contribuye al desarrollo de esta enfermedad (gluten) es ampliamente conocido (si bien otros posibles factores ambientales podrían estar influyendo [195, 196]), gran parte del componente genético se desconoce. Hasta hace poco, los únicos factores genéticos de riesgo descritos eran la combinación alélica HLA-DQ2 (*DQA1*05-DQB1*02*) y HLA-DQ8 (*DQA1*0301/DQB1*0302*), que aparecen en más del 95% de los enfermos. Sin embargo, la insuficiente contribución de estos factores al riesgo genético de la EC (aproximadamente el 50%) [67-69], además de su presencia en un cuarto de la población caucásica sana, implican la existencia de factores genéticos adicionales relevantes en esta enfermedad.

En este trabajo nos planteamos avanzar en la búsqueda de estos nuevos genes de susceptibilidad a EC con estudios de asociación siguiendo modelos caso-control y estudios familiares mediante TDT (*transmission disequilibrium test*), así como la realización de un estudio de barrido genómico colaborando con otros grupos internacionales. Para ello inicialmente analizamos polimorfismos en genes que se han visto previamente asociados a EC en estudios individuales o de barrido genómico pero cuya implicación en EC no está confirmada (*ICAM1*, *SERPINE2*, *PBX3*, *PPP6C*, *IL18RAP* y *CCR3*); en genes asociados a otras enfermedades autoinmunes, basándonos en la existencia de diversos genes de susceptibilidad compartidos por distintas enfermedades (*IL23R*, *ATG16L1*, *IRGM*, *NKX23* y *CLEC16A* (*KIAA0350)*) y en genes cuya función haga sospechar un papel importante en la aparición de EC (*IL6*, *CIITA*, *NOS2A*). Finalmente se llevó a cabo el análisis de polimorfismos de un único nucleótido (SNPs, *single nucleotide polymorphisms*) repartidos a lo largo de todo el genoma para intentar detectar nuevas señales de susceptibilidad.

Un factor genético puede ser considerado de susceptibilidad/protección al ser confirmado en muestras adicionales e independientes.

Los estudios de asociación han sido los más utilizados y con mayor éxito en la detección de factores genéticos de susceptibilidad a enfermedades complejas, destacando inicialmente la asociación de diversos genes localizados en la región *HLA* (*Human Leukocyte Antigen*) con numerosas enfermedades autoinmunes o infecciosas [197, 198], la asociación del gen *APOE* (apolipoproteína E) con la enfermedad de Alzheimer [199] o el gen *NOD2*/*CARD15* (*nucleotide-binding oligomerization domain containing* 2) con enfermedad de Crohn [200]. Sin embargo, este tipo de estudios puede presentar algunos inconvenientes, como la posibilidad de originar resultados falsos positivos (error estadístico tipo 1, rechazar la hipótesis nula de ausencia de asociación cuando es verdadera). Esto conlleva la incapacidad de replicar la asociación original en estudios posteriores, demostrando la importancia de tales replicaciones para corroborar la validez de un factor de susceptibilidad.

El polimorfismo no sinónimo G241R del gen ***ICAM1*** (*intercellular adhesión molecule* 1), sito en la región de ligamiento 19p13 (*CELIAC4* [91, 92]), fue descrito como un factor genético de elevado riesgo en población francesa [97]. ***SERPINE2*** (2q33), ***PPP6C*** y ***PBX3*** (9q34), localizados también en regiones de ligamiento a EC [84, 109], fueron destacados como posibles factores genéticos de susceptibilidad a EC en población española [121]. Ninguna de las señales de asociación en estas tres regiones fue detectada en nuestra muestra, a pesar de la elevada potencia estadística para detectar el efecto descrito. El verdadero efecto posiblemente no es tan intenso como el del estudio original, sino que podría verse inflado (*winner´s curse*) [201, 202] pero con nuestra muestra podríamos descartar efectos más modestos que los originalmente descritos con una potencia del 80% (OR=1,6, OR=0,7, OR=1,3 y OR=1,5 en *ICAM1*, *SERPINE2*, *PP6C* y *PBX3*, respectivamente), por tanto nuestros resultados parecen descartar un papel de los polimorfismos estudiados en *ICAM1*, *SERPINE2*, *PPP6C* y *PBX3* en EC.

Las regiones **2q12** (***IL18RAP*** e ***IL18R1***) y **3p21** (que contiene un "cluster" de receptores de citoquinas que incluye ***CCR1***, ***CCR2***, ***CCR3***, ***CCR5***…) fueron detectadas originalmente en un estudio GWA *(genome wide association)* en EC [132]; sin

embargo, no se vieron asociadas posteriormente en población italiana [135]. No obstante, el estudio de ambas regiones en nuestra población española mostró su implicación en EC, descartando la existencia de un efecto exclusivo en ciertas poblaciones europeas como inicialmente se había postulado. Mientras que la asociación en la región 3p21 se detecta claramente en nuestro estudio (OR=1,32, p=0,004), la señal en 2q12 evidenció la existencia de un efecto menor (OR combinada de Mantel-Haenszel=1,12, p=0,04) al descrito inicialmente (OR=1,29) [132]. Esto implica una escasa potencia estadística para detectarlo, explicando la ausencia de asociación al estudiarlo en nuestra población de forma aislada, así como en otras poblaciones [134, 135]. De hecho, nuestro estudio de 2q12 requirió la realización de un meta-análisis empleando los resultados negativos previos. Los métodos de meta-análisis evalúan y combinan los distintos efectos obtenidos en diversos estudios y son la estimación más fiable para detectar sobreestimaciones de los estudios originales [203].

La detección de factores de susceptibilidad en ciertas poblaciones y su ausencia en replicaciones posteriores podría deberse también a la existencia de una diversidad genética en la susceptibilidad a EC entre poblaciones, apoyada por las diferencias genéticas poblacionales (tanto en frecuencias como en patrones de bloques de desequilibrio) existentes incluso en la región europea [204]. Esta diversidad se ha postulado, por ejemplo, en relación al gen *MYO9B* (*myosin IXB*) [93], localizado en la región de ligamiento 19p13. Sin embargo, Lohmuller y colaboradores [205], tras el análisis de varias asociaciones genéticas, sugieren que la diversidad genética no parece explicar la gran heterogeneidad de los resultados observada en la literatura. Estudios de meta-análisis, como el realizado en relación a la región 2q12, ayudarían a descartar una posible heterogeneidad genética como la responsable de los resultados discrepantes en las distintas poblaciones europeas.

Los estudios GWA también están sujetos a la confirmación de las señales de asociación mediante replicaciones. Así el estudio mostrado en este trabajo corrobora el riesgo que ofrecen en la susceptibilidad a EC las 13 regiones previamente descritas (*RGS1*, *REL*, *IL18RAP*, *ITGA4*, *CTLA4/ICOS/CD28*, *CCR3*, *IL12A*, *LPP*, *IL2/IL21*, *TNFAIP3*, *TAGAP*, *SH2B3*, *PTPN2*) [126, 132, 133, 150], además de no manifestar resultados heterogéneos entre las distintas poblaciones analizadas, a excepción del *HLA*.

Una susceptibilidad genética común entre las distintas enfermedades autoinmunes, incluyendo celiaquía, hace sospechar mecanismos fisio-patológicos compartidos.

La co-existencia de enfermedades autoinmunes o inflamatorias dentro de la misma familia o incluso en un mismo individuo, lleva a la propuesta de la existencia de mecanismos fisio-patológicos similares en las distintas enfermedades. Ello podría deberse a factores genéticos comunes implicados en la susceptibilidad a dicho grupo de enfermedades. Esta idea ha sido tradicionalmente avalada por los estudios genéticos de la región *HLA* [206], o en regiones fuera del *HLA* como 2q33, que contiene los genes *CD28-CTLA4-ICOS* [112, 207], o 1p13, donde se localiza el gen *PTPN22* (*protein tyrosine phosphatase*, *non-receptor type 22*) [208], que se han visto asociadas a numerosas enfermedades autoinmunes.

La enfermedad inflamatoria intestinal (EII), que engloba dos formas mayoritarias que son enfermedad de Crohn y colitis ulcerosa (CU), es un desorden inflamatorio crónico intestinal que al igual que la EC se desencadena debido a una pérdida de tolerancia oral, aunque en este caso parece que es mediada por bacterias comensales que residen en el lumen intestinal (a diferencia del gluten en EC). La diabetes tipo 1 (DT1) consiste en una destrucción de las células β del páncreas que lleva a la insuficiente producción de insulina, pero también se desarrolla un proceso inflamatorio y una alteración de la permeabilidad intestinal en respuesta a antígenos de la dieta o de la flora comensal [209]. Suelen aparecer dentro de la misma familia o incluso en el mismo individuo afectado, la EC con EII [210] y con DT1 [211]. Existen pacientes celíacos que muestran manifestaciones neurológicas [212, 213] y a pesar del desconocimiento acerca de los mecanismos patogénicos implicados, también se ha observado la coexistencia de la esclerosis múltiple (EM) y EC en la misma familia [214, 215].

Las señales detectadas originalmente en estudios GWA realizados en EII, DT1 y EM, establecen firmemente la implicación de regiones que incluyen genes como ***IL23R*** (*interleukin* 23 *receptor*) (**1p31**) [123], ***ATG16L1*** (*autophagy-related 16-like 1*) (**2q37**) [216], ***IRGM*** (*immunity-related GTPase family M*) (**5q33**) y ***NKX2-3*** (*NK2 transcription factor related*, *locus 3*) (**10q24**) en EII [124] y ***CLEC16A*** (*C-type lectin*

domain family 16, *member A*) (**16p13**) en DT1 [161] y EM [162]. Al estudiar la implicación de estos genes en EC observamos diversos resultados que se discuten a continuación.

El alelo Gln381 del gen ***IL23R***, inicialmente descrito como protector frente a la enfermedad de Crohn, se ve asociado a un mayor riesgo a padecer celiaquía en nuestra muestra de población española. Esta región se ha visto recientemente ligada a EC en población finlandesa [217], confirmándose su posible implicación en esta enfermedad, a pesar de la ausencia de asociación descrita en población holandesa [218]. Sin embargo, hay que tener en cuenta que en población holandesa se observa el mismo efecto que en nuestra población (7,8% de Gln381 en celíacos vs 6,4% en controles en Holanda; 8,3% en celíacos vs 5,7% en controles en España) y, por tanto, probablemente su resultado no es significativo debido a la falta de potencia estadística. Además, un meta-análisis combinando los datos de ambos estudios produce un aumento de la significación obtenida en nuestro estudio (p=0,008; OR combinada de Mantel-Haenszel =1,35; 95% IC=1,08-1,68). Este polimorfismo Arg381Gln muestra el mismo efecto en un estudio en EM en nuestra población, pero este efecto es opuesto al descrito en otras enfermedades autoinmunes como EII [123], psoriasis [219] y espondilitis anquilosante [220]. Esta diferencia podría sugerir la existencia de efectos moleculares opuestos en EC y EM con respecto a las otras patologías mencionadas. No obstante, este polimorfismo podría ser un marcador cuyas variantes alélicas estuvieran en desequilibrio de ligamiento parcial o incompleto con una variante etiológica distinta en cada enfermedad. Dependiendo del grado de correlación entre los alelos del polimorfismo Arg381Gln y la variante causal en cada enfermedad, existirá también una mayor o menor correlación entre ambas variantes causales (Figura 7). Es importante resaltar que el papel funcional de ese cambio amino-acídico aún no se conoce.

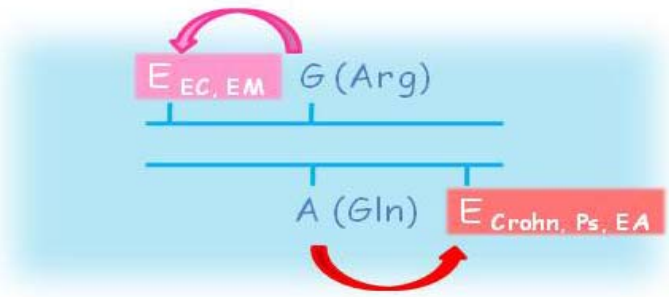

Figura 7: Representación de una posible explicación de la asociación observada con Arg381Gln en las distintas enfermedades.
$E_{EC, EM}$: factor etiológico en enfermedad celíaca y esclerosis múltiple.
$E_{Crohn, Ps, EA}$: factor etiológico de enfermedad inflamatoria intestinal (EII), psoriasis (Ps) y espondilitis anquilosante (EA).

Existen otros ejemplos en la literatura en los que el mismo polimorfismo se ha visto asociado a diferentes enfermedades con efectos opuestos, como por ejemplo las regiones 4q27 (*IL2-IL21*) y 6q25 (*TAGAP*), asociadas a DT1 [127] y a celiaquía.

En general, el hecho de que el fenotipo de un individuo dependa de la combinación específica de alelos en dos o más *loci*, podría sugerir que se debiera a la interacción o epistasis entre dichos alelos. La epistasis podría entonces manifestar su efecto más claramente en fenotipos que suponen patologías [221, 222]. Encontramos un ejemplo con el alelo *HLA-DRB5*0101* que parece modular la severidad de la EM y se mantiene en desequilibrio de ligamiento por selección natural positiva con el alelo *HLA-DRB1*1501*, uno de los principales factores genéticos de riesgo que determinan la susceptibilidad a EM [223]. En psoriasis, se observa un aumento del riesgo a padecer la enfermedad al combinar el efecto de ***IL23R*** con ***IL12B*** (codifica la subunidad p40, compartida por las citoquinas heterodiméricas IL-12 e IL-23) [219], lo que sugiere la presencia de epistasis entre genes que participan en una ruta molecular determinada, entendiendo epistasis como una desviación del efecto alélico aditivo esperado [224]. Aunque dicha interacción no parece estar implicada en la susceptibilidad a EC, la búsqueda de interacciones gen-gen o gen-ambiente puede ser una estrategia a tener en cuenta para la comprensión de la base genética de la enfermedad.

Los polimorfismos estudiados en las regiones ***ATG16L1*** (**2q37**), ***IRGM*** (**5q33**), ***NKX2-3*** (**10q24**) y ***CLEC16A*** (**16p13**), anteriormente citadas, no parecen estar implicadas en la susceptibilidad a EC en nuestra población. Nuestro estudio ofrece resultados negativos a pesar de alcanzar la potencia estadística suficiente para detectar el riesgo descrito en EII, DT1 y EM (potencia ≥ 80% para detectar ORs comprendidas entre 1,22 y 1,44 dependiendo del SNP considerado). Sin embargo, polimorfismos que pueden resultar buenos marcadores de susceptibilidad a una enfermedad en una población pueden no serlo en otras debido a diferentes patrones de LD, de manera que el estudio de haplotipos podría ayudar a la detección de señales de susceptibilidad diferentes a las descritas. Pese a ello, confirmamos que los haplotipos formados por los SNPs estudiados en los genes ***IRGM*** (**5q33**), ***NKX2-3*** (**10q24**) y ***CLEC16A*** (**16p13**) tampoco muestran asociación a EC en nuestra población.

Debido a la distribución del genoma en bloques haplotípicos de alto desequilibrio de ligamiento (*D´*=1) [103], las señales detectadas en los estudios GWA dentro de un mismo bloque podrían marcar una única variante etiológica o varias independientes "ocultas" tras el desequilibrio. Los genes adyacentes ***CIITA*** y ***CLEC16A*** se encuentran en la región **16p13** y aunque no parecen estar incluidos en el mismo bloque de desequilibrio, los marcadores de ambos genes asociados a DT1 y EM podrían estar apuntando la misma variante etiológica. Hasta la era de los GWAS se adjudicaba al gen *CIITA* el riesgo de susceptibilidad descrito en DT1 [225] y EM [120], aunque con numerosas discrepancias. Sin embargo, los estudios GWA de DT1 [161] y EM [162] detectaron señales de susceptibilidad en el gen *CLEC16A*. Las variantes alélicas de riesgo detectadas en dichas enfermedades se encuentran en suficiente desequilibrio de ligamiento, como muestro en este trabajo, como para que resulte preciso un análisis condicionado o estratificado para averiguar la independencia de dichas señales o para revelar la presencia de factores de susceptibilidad enmascarados. Aunque el análisis estratificado en nuestra muestra no evidencia la presencia de factores genéticos ocultos en esta región asociados a EC, parece necesaria una mayor clarificación en DT1 y EM [226].

Si bien las regiones *ATG16L1* (2q37), *IRGM* (5q33), *NKX2-3* (10q24) y *CLEC16A* (16p13), previamente asociadas a otras enfermedades de carácter autoinmune, no parecen estar implicadas en la susceptibilidad a EC, estas

enfermedades parecen compartir factores genéticos con EC de forma más frecuente a la esperada (Figura 8). Por tanto, el análisis de genes con confirmada asociación a otras enfermedades autoinmunes o inflamatorias, parece ser una aproximación válida para la búsqueda de factores genéticos en EC.

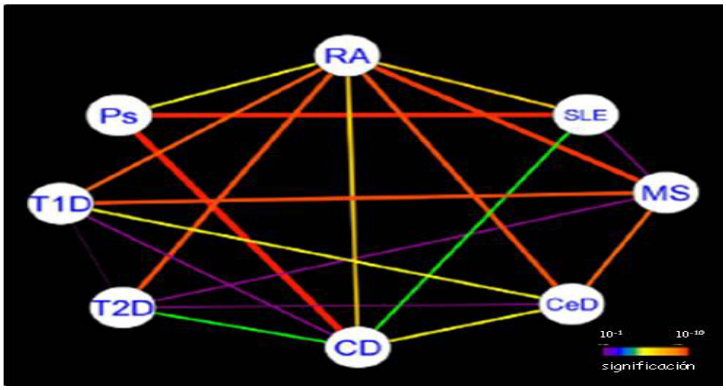

Figura 8: Red de relaciones entre enfermedades. Los colores muestran la significación del exceso de genes compartidos. (Figura obtenida de Baranzini et al. [227]).
RA, *rheumatoid arthritis*; SLE, *systemic lupus erythematosus*; MS, *multiple sclerosis*; CeD, *celiac disease*; CD, *Crohn´s disease*; T2D, *type 2 diabetes*; T1D, *type 1 diabetes*; Ps, *psoriasis*.

Sin embargo hay que tener en cuenta que en el estudio de colaboración que muestro en este trabajo se observa como 18 de las 27 regiones que ofrecen un riesgo a padecer celiaquía se encuentran también asociadas a otras enfermedades inflamatorias o inmunológicas, pero en tan sólo 3 regiones la asociación se detecta con el mismo SNP o uno equivalente (r^2>0,8). De esta manera la detección de señales comunes exigiría un análisis más exhaustivo de la región, puesto que la asociación observada puede ser consecuencia del riesgo conferido por una misma variante pero con distinto efecto o incluso de distintas variantes en la misma región.

La base genética de los procesos patológicos que se desarrollan a nivel local en celiaquía no se encuentra completamente resuelta.

Los análisis del perfil de expresión [228-230] o los ensayos de análisis de citoquinas [31, 231] u otros componentes, contribuyen a la comprensión de los mecanismos funcionales que llevan a la patogénesis de la EC. La variación de los niveles de ciertas moléculas ha supuesto una de las estrategias más utilizadas para apuntar a una posible implicación genética en la EC. Así, un aumento en la expresión de la citoquina IL-6 [32, 232] y del enzima iNOS (óxido nítrico sintasa inducible) [233] en pacientes celíacos, podría sugerir la existencia de variaciones genéticas en regiones reguladoras de la transcripción de sus genes codificantes ***IL6*** y ***NOS2A***, respectivamente, que posiblemente modificarían el riesgo a padecer celiaquía.

La variante -174G/C de la región promotora del gen ***IL6*** (7p15), junto con otros elementos reguladores presentes en el mismo haplotipo, parece que modula los niveles de expresión de la IL-6 [234, 235]. A pesar de la complejidad en la regulación de su expresión, así como de la especificidad de dicha modulación en función del tipo celular [235], los estrógenos parecen estar implicados [236], pudiéndose observar un comportamiento diferencial entre sexos [237]. El polimorfismo *IL6* -174G/C no mostró asociación a EC en estudios previos [238]; sin embargo, el aumento de riesgo que ofrece el alelo *IL6* -174C específicamente en niñas celíacas en nuestra población, podría explicar la ausencia de detección previa. Este efecto también se observa en nuestro estudio familiar independiente, descartando la presencia de un posible falso positivo por estratificación poblacional. Aunque la implicación funcional de esta variante alélica en la susceptibilidad de la enfermedad se desconoce, este efecto apoyaría la diferencia genética entre sexos esperada debido a la distinta prevalencia observada tanto en EC como en otras enfermedades autoinmunes. El efecto de los estrógenos en el desarrollo y función de las células inmunológicas, en la producción de citoquinas y quimioquinas o en la migración celular a las regiones de inflamación, puede afectar a la patogénesis de enfermedades crónicas autoinmunes o inflamatorias y puede contribuir a la mayor incidencia de ciertas enfermedades autoinmunes en mujeres [239].

Por otro lado, la variante Asp358Ala del gen ***IL6R*** (1q21), gen que codifica la subunidad IL-6Rα del receptor dimérico de la citoquina IL-6, se ha visto fuertemente

relacionado con los niveles plasmáticos de IL-6 [240, 241], además de controlar los niveles de la forma soluble del receptor (sIL-6Rα) [242]. Aunque dicha variante no muestra una evidencia clara de riesgo en EC, otros mecanismos podrían distorsionar la asociación de *IL6R*, como la posible interacción con la variante asociada a EC en el gen *IL6*. Sin embargo, tampoco se observa efecto en el gen *IL6R* al estratificar por las variantes estudiadas del gen *IL6*.

La regulación del gen ***NOS2A*** a nivel transcripcional se ha considerado el mecanismo de regulación más importante en la expresión de iNOS, detectándose en el promotor numerosas secuencias de unión de factores de transcripción inducidos, entre otros, por citoquinas proinflamatorias [187]. La implicación de las distintas variantes localizadas en el promotor del gen que pudieran estar influyendo en esta regulación ha sido ampliamente estudiada. El alelo $(CCTTT)_{14}$ del microsatélite pentanucleotídico que estudiamos en este trabajo, parece promover la expresión de *NOS2A* de forma más eficaz tras la estimulación con la citoquina proinflamatoria IL-1β [118]. A su vez, la variante (TAAA)ins del microsatélite bi-alélico tetranucleotídico, también analizado en este trabajo, parece inducir una mayor expresión de iNOS [243]. Sin embargo, ningún alelo de ambos microsatélites parece mostrar un efecto en la susceptibilidad a EC en nuestra muestra de población española, aunque dichas variantes se han visto asociadas a otras enfermedades como DT1 o EM [118, 244] entre otras.

Otra de las ventajas de estudiar haplotipos, además de cubrir una mayor variabilidad génica, es establecer la fase de las distintas variantes alélicas (presencia en *cis* o en *trans* de los alelos de cada locus). Su presencia en *cis* implica desequilibrio de ligamiento entre las distintas variantes, que puede ser consecuencia de una implicación funcional como podría ser la participación en la regulación de la transcripción. La presencia de dos microsatélites, (TAAA)ins/del y $(CCTTT)_n$, en la región promotora del gen *NOS2A*, el segundo de ellos con una elevada tasa de mutación, podría hacer intuir una muy elevada diversidad haplotípica por la rotura del desequilibrio de ligamiento y por tanto descartar la idea de una implicación de algún haplotipo específico en la transcripción [245]. Sin embargo, como muestro en este trabajo, determinados alelos del microsatélite pentanucleotídico (que ofrece la mayor tasa de mutación) se encuentran formando haplotipos con otros alelos de otros marcadores de la misma región, pudiéndose observar además la presencia en el mismo

haplotipo de las variantes (TAAA)ins y $(CCTTT)_{14}$, anteriormente citadas como variantes inductoras de la expresión de *NOS2A*. No obstante, por el análisis de los haplotipos extendidos parece que este gen no modula el riesgo a padecer EC en nuestra muestra.

La ausencia de efecto de la región promotora del gen *NOS2A* sobre el riesgo a padecer EC, observada a pesar del aumento en los niveles de expresión detectados, podría llevar a distintas conclusiones. Debido a que la regulación de la transcripción es muy compleja, distintos mecanismos moduladores, aparte de los que actúan en la región promotora, pueden estar interviniendo en la expresión, bien a nivel del mismo gen pero variabilidad no recogida en nuestro trabajo, o bien controlada por otros factores externos al gen. Así mismo, también hay que considerar que la variación en los niveles de expresión detectados y asociados a una determinada patología, pueden ser secundarios a la progresión de la enfermedad y no causa de la misma.

El conocimiento de la posible implicación funcional de los distintos factores genéticos de susceptibilidad a EC detectados, supondría el siguiente paso a seguir.

La identificación de las variantes genéticas etiológicas implicadas en una determinada enfermedad, como es la EC, es un proceso complejo, ya que las señales de susceptibilidad detectadas en un punto del genoma pueden corresponder a diversas variantes genéticamente equivalentes que pueden estar situadas en distintas posiciones del gen o incluso estar en distintos genes. Si la señal detectada se encuentra ubicada en regiones reguladoras o regiones codificantes de un gen, podría considerarse que esa variante pudiera tener una implicación funcional en el desarrollo de la patología. Así el polimorfismo -174G/C localizado en la región promotora del gen *IL6*, o bien la variante de la región codificante del gen *IL23R*, Arg381Gln, cuyos alelos minoritarios muestran un mayor riesgo a padecer celiaquía, podrían tener una importancia en la fisio-patogenia de la EC, aunque de momento se desconoce. Sin embargo, en la gran mayoría de los casos, se hace casi imposible atribuir a una señal una implicación funcional, bien porque se encuentra en regiones no codificantes de genes, en los denominados desiertos génicos o en regiones intergénicas. En estos casos, se deduce

de nuestro último trabajo, que probablemente muchas de estas variantes estén afectando a los niveles de expresión. De esta manera, la señal de susceptibilidad a EC detectada en la región 2q12 se encuentra comprendida entre los genes *IL18R1* e *IL18RAP*, genes que codifican las subunidades IL-18Rα e IL-18Rβ, respectivamente, y cualquiera de ellos, sino ambos, se podría pensar que estar implicados en la patogenia de la enfermedad. Sin embargo, el estudio de expresión parece indicar un papel de *IL18RAP* por ser su expresión la que se ve correlacionada.

Por otro lado, algunas de las señales detectadas se podrían considerar marcadores de susceptibilidad que presentan una elevada correlación con la variante causal, pero cuya presencia en un bloque de elevado desequilibrio de ligamiento hace complicada la identificación de dicha variante causal. Así, la variante de riesgo a EC de la región 3p21 se encuentra entre los genes *CCR3* y *CCR2*, ambos en un bloque de desequilibrio que comprende un "cluster" de genes de receptores de citoquinas que también incluye *CCR1*, *CCR5* y *CCRL2*, entre otros, y todos ellos con una posible importancia en el desarrollo de la enfermedad.

Los principales factores genéticos de susceptibilidad a EC, *DQA1*05-DQB1*02* y *DQA1*0301/DQB1*0302*, que codifican las moléculas HLA-DQ2 y DQ8, respectivamente, se sabe que juegan un papel en la presentación de péptidos de gliadina, determinando así la magnitud de la respuesta específica de células T y, por tanto, el desarrollo de la enfermedad. Aunque no existe una implicación funcional tan clara para los restantes factores genéticos de susceptibilidad a EC, la mayoría de estos genes codifican proteínas a las que se les atribuye una función inmunológica, por lo que podrían estar interviniendo en la patogenia de la EC. Sin embargo, hay que resaltar que el papel específico que están desempeñando en la patogenia las variantes alélicas que influyen en la susceptibilidad a EC, aún no se conoce.

Los efectos sugeridos por los estudios individuales desarrollados a lo largo de este trabajo se resumen en la Figura 9.

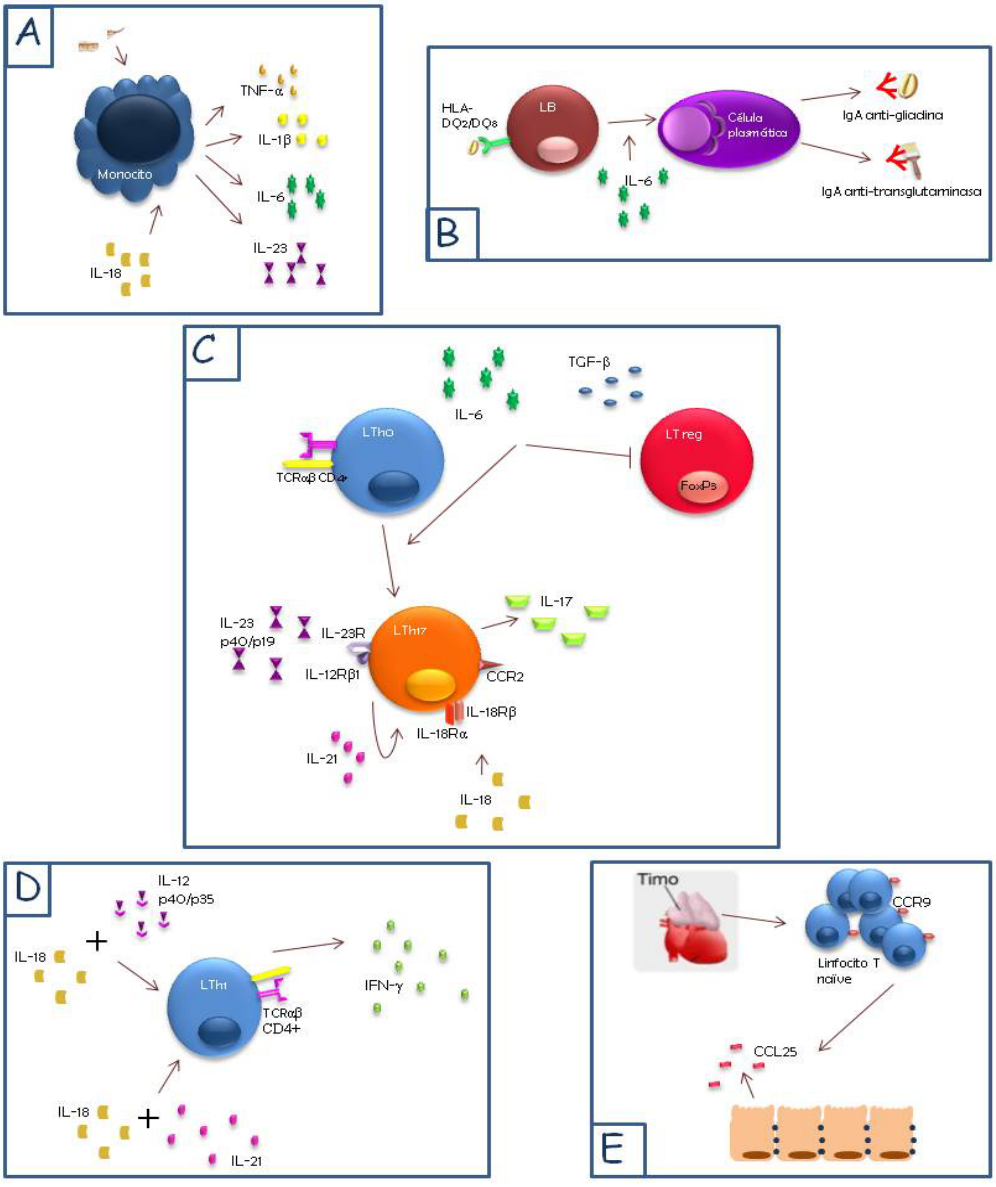

Figura 9: Posible implicación funcional en la inmuno-patogenia de la EC de las moléculas codificadas por los genes que muestran riesgo en nuestro trabajo. A. Secreción de citoquinas pro-inflamatorias tras estimulación con gliadina. B. Producción de anticuerpos favorecida por la IL-6. C. Inducción y mantenimiento de los linfocitos Th17. D. Producción de IFN-γ favorecida por la IL-18. E. Migración de los linfocitos T a la mucosa intestinal.
TNF-α, *tumor necrosis factor* α; IL- interleuquina; HLA, H*uman Leukocyte Antigen*; Ig, inmunoglobulina; TGF-β, *transforming growth factor β*; CCR, *chemokine* C-C *motif receptor*; IFN-γ, *γ interferon*; CCL, *chemokine* C-C *motif ligand*.

En la patogenia de la EC se desarrollan distintos mecanismos de inmunidad innata y adaptativa que se suceden y se solapan, llevando a la cronicidad del proceso inflamatorio. La detección de una producción de citoquinas pro-inflamatorias, a partir de monocitos aislados de sangre periférica de pacientes celíacos, como IL-1β, TNF-α, IL-6 y de IL-23, tras la estimulación con fragmentos peptídicos de gliadina [246], podría sugerir la existencia de una activación de macrófagos localizados en la lámina *propria* del intestino, desencadenando una respuesta inicial inespecífica del sistema inmunológico cuyos mecanismos moleculares implicados aún no son del todo conocidos. Esta hipótesis se ve apoyada por la detección de un aumento de los niveles de dichas citoquinas pro-inflamatorias en mucosa intestinal de pacientes celíacos activos [247, 248], además de detectarse en suero [232, 249]. Además, los niveles proteicos de la forma madura de la IL-18 se encuentran elevados en la mucosa intestinal de pacientes celíacos [35], favoreciendo la respuesta inmune innata al inducir la expresión de citoquinas pro-inflamatorias [250] (Figura 9A).

La IL-6 está implicada en multitud de procesos. En la mucosa intestinal desempeña diversas funciones y por ello podría actuar a distintos niveles en la patogenia de la enfermedad. Por un lado interviene en la inducción de la maduración final de células B a células plasmáticas [251], lo que podría favorecer la producción de anticuerpos anti-gliadina y anti-TG2. Esto parece confirmarse por la detección de una correlación positiva entre los títulos de anticuerpos EMA-IgA (IgA anti-endomisio) y los niveles de IL-6 en suero de pacientes celíacos [232] (Figura 9B). Así mismo, la IL-6 también podría regular el balance de células T reguladoras y células Th17, con lo cual se interrumpiría el proceso de tolerancia oral con la consecuente pérdida de la homeostasis intestinal [252]. La IL-6 inhibe el efecto inductor del TGF-β sobre las células T reguladoras [253] y a su vez, en presencia de TGF-β y otras citoquinas pro-inflamatorias, favorece la diferenciación de células T naïve a células Th17, caracterizadas por la producción de IL-17, y puede desencadenar así un proceso autoinmune e inflamatorio [254, 255] (Figura 9C). En el proceso de diferenciación de las células Th17, la IL-6 también promueve la expresión de **IL-23R** en superficie, por tanto aumenta la capacidad de respuesta a la IL-23, lo que estimula la expresión de IL-17. Por tanto se considera esencial en el mantenimiento del fenotipo Th17 [256]. La implicación del eje IL-23/Th17 en la inducción de diversas enfermedades inflamatorias, incluyendo las enfermedades autoinmunes, es un hecho que se está

estudiando ampliamente. En celiaquía se observa un aumento en los niveles de expresión de *IL23A* (que codifica la subunidad p19 de la IL-23) e *IL17A* (que codifica la IL-17) en biopsias intestinales [257]. En células dendríticas de mucosa de pacientes celíacos se ha detectado un aumento de los transcritos de la subunidad p19 de la IL-23 [33], pero una nula producción de la subunidad p40 IL-12/IL-23 [32, 33]. Esto promueve la idea de un escaso papel de la IL-23 en la patogenia de la EC, sin embargo, otras células de la respuesta innata, como los macrófagos y monocitos, pueden estar interviniendo en la producción de dicha citoquina [246] (Figura 9C).

A su vez, las células Th17 secretan IL-21, que actúa de forma autocrina amplificando su diferenciación tras la inducción con IL-6 [258]. Este efecto podría implicar un papel de la IL-21 en la patogenia de la EC, corroborado tanto por el aumento de sus niveles en mucosa intestinal de pacientes celíacos [36] como por la fuerte evidencia de asociación de la región cromosómica 4q27 con EC [126] (Figura 9C).

Se ha visto que la IL-18, sinérgicamente con la IL-23, amplifica la producción de la IL-17 por parte de las células Th17 [259]. Sin embargo, parece que la responsable de esta acción no es tanto la producción de IL-18, como el correcto ensamblaje de la subunidad IL-18Rα de su receptor [260]. Dicha subunidad se encuentra codificada en la región cromosómica 2q12, y se ve asociada a EC, como muestro en este trabajo. Además, se ha detectado un aumento en la expresión de ambas cadenas del receptor de la IL-18 en la mucosa intestinal de pacientes celíacos [35] (Figura 9C).

Las células Th17 se encuentran definidas por receptores de quimioquinas, algunos de los cuales se encuentran codificados en la región 3p21, que muestra una asociación a EC confirmada en este trabajo. Tanto la co-expresión de receptores CCR4 (3p24) y CCR6 (6q27) [261], como la expresión de **CCR2** (3p21) en ausencia de **CCR5** (3p21) [262], marcan el fenotipo Th17 (Figura 9C).

La IL-18 y la IL-21 además presentan un papel en la respuesta Th1, tradicionalmente considerada el tipo de respuesta dominante en EC. Actualmente se sabe que existe una inducción de la producción de IFN-γ al combinar IL-21 e IL-18 o IL-15 [263], citoquinas con una elevada expresión en la mucosa intestinal. Hasta hace poco la secreción de IFN-γ se creía que ocurría tan sólo tras la estimulación conjunta de la IL-18 con la IL-12 [264], esta última ausente en biopsias intestinales [32]. Sin

embargo, la IL-12 podría tener una posible implicación en la patogenia de la enfermedad debido a la señal de susceptibilidad observada en la región cromosómica 3q25 [132], donde se encuentra localizado el gen *IL12A* que codifica la subunidad p35 del complejo heterodimérico p40/p35 que consitituye la IL-12 (Figura 9D).

El receptor de quimioquinas **CCR9**, codificado en la región cromosómica 3p21 de susceptibilidad a EC, se considera un receptor que favorece la migración de linfocitos TCRγδ del timo y otros linfocitos T (CD4+ y CD8+) hacia la mucosa intestinal en respuesta a CCL25, expresado por las células epiteliales intestinales [265, 266]. A su vez, parece que el nivel de expresión de CCR9 en células T es modulado por el grado de activación de estas células, observándose una disminución de su expresión en mucosa de pacientes celíacos no tratados pero también en tratados, lo que sugiere que la activación de células T es una alteración duradera [267] (Figura 9E).

Aparte de todos estos factores genéticos que incrementan el riesgo a EC, como consecuencia de los estudios de barrido genómico o sus *follow-up* se han descrito otras regiones de susceptibilidad en las que los genes candidatos más probables codifican proteínas con una posible implicación funcional inmunológica en la patogenia de la EC. Las regiones asociadas con los posibles genes candidatos y su posible función se detallan en la Tabla 3.

Tabla 3: Otros *loci* descritos de susceptibilidad a EC. Tabla adaptada de Jabri et al. [268].

Locus	Gen candidato	Implicación funcional	Referencia
2q33	*CTLA4/ICOS*	CTLA4 es un receptor de células T para CD80/CD86 y regula negativamente la activación de células T. ICOS es un receptor de células T activadas cuya unión a su ligando LICOS, favorece la proliferación celular.	[108]
1q31	*RGS1*	RGS1 está implicada en señalización celular y se expresa en LIE.	[132]
3q28	*LPP*	Desconocida.	[132]
6q25	*TAGAP*	TAGAP se expresa en células T activadas y es importante en regular los cambios del citoesqueleto.	[132]
12q23	*SH2B3*	Codifica la proteína LNK (proteína adaptadora de linfocitos) implicada en señalización de linfocitos.	[132]
18p11	*PTPN2*	Tirosín-Fosfatasa considerada regulador negativo de la inflamación.	[127]
6q23	*TNFAIP3*	Proteína en "dedo de zinc" que inhibe la actividad NF-κB y la muerte celular programada mediada por TNF.	[150]
2q31	*ITGA4*	Codifica la subunidad α4, formando heterodímeros con β1 y β7, promoviendo la adhesión y reclutamiento de linfocitos al intestino, al unirse a las fibronectinas y MAdCAM-1, respectivamente.	[133]
2p13	*REL*	Componente del complejo de transcripción NF-κB.	[150]

LIE, linfocitos intraepiteliales; CTLA4, *cytotoxic T lymphocyte antigen 4*; ICOS, *inducible T-cell co-stimulator*; LICOS, ICOS *ligand*; RGS1, *regulator of G protein signalling* 1; LPP, LIM *domain containing preferred translocation partner in lipoma*; TAGAP, *T cell activation RhoGTPase activating protein*; SH2B3, *SH2B adaptor protein 3; PTPN2, protein tyrosine phosphatase, non-receptor type 2;* TNFAIP3, *tumour necrosis factor α-induced protein; ITGA4, integrin alpha 4;* MAdCAM-1, *mucosal addressin-cell adhesión molecule; REL, v-rel reticuloendotheliosis viral oncogene homolog.*

Los genes incluidos tanto en estas regiones como en las nuevas apuntadas en el estudio de colaboración que muestro en este trabajo, también sugieren una relevancia inmunológica a nivel del desarrollo de células T en el timo, detección de RNA viral por el sistema de inmunidad innata o coestimulación de linfocitos T y B.

Los diversos genes descritos en la susceptibilidad a la EC revelan una gran variedad de mecanismos patológicos que conducen o se desarrollan en la enfermedad, aunque hay que tener en cuenta que los genes identificados hasta la fecha no explican todo el componente genético de la enfermedad [157]. El descubrimiento de la implicación funcional precisa de todos los factores genéticos en la patogenia de la EC,

sería un siguiente paso a esclarecer. No obstante, estos avances en la base genética permiten ampliar el conocimiento sobre procesos inmunológicos relevantes en la enfermedad y proporcionan pistas importantes sobre nuevos genes objeto de estudio, que podrían ser aquellos que codifican moléculas implicadas en estas rutas inmuno-patogénicas.

CONCLUSIONES

La enfermedad celíaca (EC) presenta un elevado componente genético de carácter complejo en el que la presencia de múltiples factores genéticos, así como la interacción entre ellos o con componentes ambientales, determina la predisposición a la enfermedad tras la ingesta de gluten. De esta manera, el estudio de su base genética es un campo muy amplio que aún no está completamente resuelto. Los avances producidos en este sentido con este trabajo permiten extraer las siguientes conclusiones:

1. Los efectos de susceptibilidad/protección previamente publicados en los genes *ICAM1* (19p13), *SERPINE2* (2q33), *PPP6C* (9q33) y *PBX3* (9q33), todos ellos localizados en regiones previamente ligadas a EC, no son replicados en nuestra muestra de población española, lo que parece descartar su implicación en la enfermedad.

2. Las regiones 2q12 (*IL18RAP*) y 3p21 (*CCR3*), cuya asociación fue identificada en el estudio de barrido genómico de celiaquía, pero no fue corroborada en estudios posteriores, muestran riesgo a padecer celiaquía en nuestra población. En 2q12 se observa un efecto de asociación débil, lo que genera limitaciones de potencia estadística para su detección.

3. La variante Gln381 del gen *IL23R* aumenta la susceptibilidad a EC, mostrando un efecto opuesto al descrito en otras enfermedades autoinmunes. Este efecto no parece verse afectado por la interacción con polimorfismos en el gen *IL12B*, a diferencia de lo observado en psoriasis.

4. Las señales de susceptibilidad a enfermedad de Crohn en los genes *NKX2-3*, *ATG16L1* e *IRGM*, no muestran evidencia de asociación a EC.

5. Polimorfismos en los genes *CIITA* y *CLEC16A*, localizados en la región 16p13, e implicados en otras enfermedades autoinmunes, no modifican el riesgo a padecer celiaquía.

6. El alelo minoritario del polimorfismo funcional del promotor del gen *IL6* -174G/C, aumenta el riesgo a padecer celiaquía en niñas.

7. Variaciones en la frecuencia de polimorfismos en la región promotora del gen *NOS2A* no afectan a la susceptibilidad a padecer EC.

8. Se confirman las señales de susceptibilidad previamente descritas en los GWAS publicados y se apuntan 26 regiones nuevas, la mayoría incluyendo genes con una implicación funcional inmunológica.

9. Los genes asociados a EC sugieren diversas alteraciones de la respuesta inmunológica que pueden contribuir al desarrollo de la EC y cuya función exacta en el proceso etiológico deberá esclarecerse.

10. El *HLA* continúa siendo el principal factor genético de susceptibilidad a EC y el único con una función claramente establecida en el proceso patológico.

BIBLIOGRAFÍA

1. Catassi, C. et al. Coeliac disease in the year 2000: exploring the iceberg. *Lancet* **343**, 200-3 (1994).
2. Dube, C. et al. The prevalence of celiac disease in average-risk and at-risk Western European populations: a systematic review. *Gastroenterology* **128**, S57-67 (2005).
3. Hoffenberg, E.J. et al. A prospective study of the incidence of childhood celiac disease. *J Pediatr* **143**, 308-14 (2003).
4. Cummins, A.G. & Roberts-Thomson, I.C. Prevalence of celiac disease in the Asia-Pacific region. *J Gastroenterol Hepatol* **24**, 1347-51 (2009).
5. van Heel, D.A. & West, J. Recent advances in coeliac disease. *Gut* **55**, 1037-46 (2006).
6. Peters, U., Schneeweiss, S., Trautwein, E.A. & Erbersdobler, H.F. A case-control study of the effect of infant feeding on celiac disease. *Ann Nutr Metab* **45**, 135-42 (2001).
7. Ivarsson, A., Persson, L.A., Nystrom, L. & Hernell, O. The Swedish coeliac disease epidemic with a prevailing twofold higher risk in girls compared to boys may reflect gender specific risk factors. *Eur J Epidemiol* **18**, 677-84 (2003).
8. Nash, S. Does exclusive breast-feeding reduce the risk of coeliac disease in children? *Br J Community Nurs* **8**, 127-32 (2003).
9. Whitacre, C.C. Sex differences in autoimmune disease. *Nat Immunol* **2**, 777-80 (2001).
10. Green, P.H.R. et al. Characteristics of adult celiac disease in the USA: results of a national survey. *Am J Gastroenterol* **96**, 126-31 (2001).
11. Steens, R.F. et al. A national prospective study on childhood celiac disease in the Netherlands 1993-2000: an increasing recognition and a changing clinical picture. *J Pediatr* **147**, 239-43 (2005).
12. Fasano, A. et al. Prevalence of celiac disease in at-risk and not-at-risk groups in the United States: a large multicenter study. *Arch Intern Med* **163**, 286-92 (2003).
13. Lo, W., Sano, K., Lebwohl, B., Diamond, B. & Green, P.H. Changing presentation of adult celiac disease. *Dig Dis Sci* **48**, 395-8 (2003).

14. Marsh, M.N. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology* **102**, 330-54 (1992).
15. Oberhuber, G., Granditsch, G. & Vogelsang, H. The histopathology of coeliac disease: time for a standardized report scheme for pathologists. *Eur J Gastroenterol Hepatol* **11**, 1185-94 (1999).
16. Corazza, G.R. et al. Comparison of the interobserver reproducibility with different histologic criteria used in celiac disease. *Clin Gastroenterol Hepatol* **5**, 838-43 (2007).
17. Kutlu, T. et al. Numbers of T cell receptor (TCR) alpha beta+ but not of TcR gamma delta+ intraepithelial lymphocytes correlate with the grade of villous atrophy in coeliac patients on a long term normal diet. *Gut* **34**, 208-14 (1993).
18. West, J., Logan, R.F., Hill, P.G. & Khaw, K.T. The iceberg of celiac disease: what is below the waterline? *Clin Gastroenterol Hepatol* **5**, 59-62 (2007).
19. Casswall, T.H., Papadogiannakis, N., Ghazi, S. & Nemeth, A. Severe liver damage associated with celiac disease: findings in six toddler-aged girls. *Eur J Gastroenterol Hepatol* **21**, 452-9 (2009).
20. Bianchi, M.L. & Bardella, M.T. Bone in celiac disease. *Osteoporos Int* **19**, 1705-16 (2008).
21. Catassi, C., Bearzi, I. & Holmes, G.K. Association of celiac disease and intestinal lymphomas and other cancers. *Gastroenterology* **128**, S79-86 (2005).
22. Ryan, B.M. & Kelleher, D. Refractory celiac disease. *Gastroenterology* **119**, 243-51 (2000).
23. Al-Toma, A., Verbeek, W.H. & Mulder, C.J. Update on the management of refractory coeliac disease. *J Gastrointestin Liver Dis* **16**, 57-63 (2007).
24. Brar, P. et al. Lack of correlation of degree of villous atrophy with severity of clinical presentation of coeliac disease. *Dig Liver Dis* **39**, 26-9; discussion 30-2 (2007).
25. Fasano, A. et al. Zonulin, a newly discovered modulator of intestinal permeability, and its expression in coeliac disease. *Lancet* **355**, 1518-9 (2000).
26. Lammers, K.M. et al. Gliadin induces an increase in intestinal permeability and zonulin release by binding to the chemokine receptor CXCR3. *Gastroenterology* **135**, 194-204 e3 (2008).

27. Schumann, M. et al. Mechanisms of epithelial translocation of the alpha(2)-gliadin-33mer in coeliac sprue. *Gut* **57**, 747-54 (2008).
28. Matysiak-Budnik, T. et al. Secretory IgA mediates retrotranscytosis of intact gliadin peptides via the transferrin receptor in celiac disease. *J Exp Med* **205**, 143-54 (2008).
29. Maiuri, L. et al. Association between innate response to gliadin and activation of pathogenic T cells in coeliac disease. *Lancet* **362**, 30-7 (2003).
30. Monteleone, I. et al. Regulation of the T helper cell type 1 transcription factor T-bet in coeliac disease mucosa. *Gut* **53**, 1090-5 (2004).
31. Nilsen, E.M. et al. Gluten specific, HLA-DQ restricted T cells from coeliac mucosa produce cytokines with Th1 or Th0 profile dominated by interferon gamma. *Gut* **37**, 766-76 (1995).
32. Nilsen, E.M. et al. Gluten induces an intestinal cytokine response strongly dominated by interferon gamma in patients with celiac disease. *Gastroenterology* **115**, 551-63 (1998).
33. Di Sabatino, A. et al. Evidence for the role of interferon-alfa production by dendritic cells in the Th1 response in celiac disease. *Gastroenterology* **133**, 1175-87 (2007).
34. Monteleone, G. et al. Role of interferon alpha in promoting T helper cell type 1 responses in the small intestine in coeliac disease. *Gut* **48**, 425-9 (2001).
35. Salvati, V.M. et al. Interleukin 18 and associated markers of T helper cell type 1 activity in coeliac disease. *Gut* **50**, 186-90 (2002).
36. Fina, D. et al. Interleukin 21 contributes to the mucosal T helper cell type 1 response in coeliac disease. *Gut* **57**, 887-92 (2008).
37. Esposito, C. et al. Anti-tissue transglutaminase antibodies from coeliac patients inhibit transglutaminase activity both in vitro and in situ. *Gut* **51**, 177-81 (2002).
38. Meresse, B., Ripoche, J., Heyman, M. & Cerf-Bensussan, N. Celiac disease: from oral tolerance to intestinal inflammation, autoimmunity and lymphomagenesis. *Mucosal Immunol* **2**, 8-23 (2009).
39. Gianfrani, C. et al. Celiac disease association with CD8+ T cell responses: identification of a novel gliadin-derived HLA-A2-restricted epitope. *J Immunol* **170**, 2719-26 (2003).

40. Mazzarella, G. et al. Gliadin activates HLA class I-restricted CD8+ T cells in celiac disease intestinal mucosa and induces the enterocyte apoptosis. *Gastroenterology* **134**, 1017-27 (2008).
41. Hue, S. et al. A direct role for NKG2D/MICA interaction in villous atrophy during celiac disease. *Immunity* **21**, 367-77 (2004).
42. Jabri, B. et al. Selective expansion of intraepithelial lymphocytes expressing the HLA-E-specific natural killer receptor CD94 in celiac disease. *Gastroenterology* **118**, 867-79 (2000).
43. Meresse, B. et al. Reprogramming of CTLs into natural killer-like cells in celiac disease. *J Exp Med* **203**, 1343-55 (2006).
44. Meresse, B. et al. Coordinated induction by IL15 of a TCR-independent NKG2D signaling pathway converts CTL into lymphokine-activated killer cells in celiac disease. *Immunity* **21**, 357-66 (2004).
45. Zeng, R. et al. Synergy of IL-21 and IL-15 in regulating CD8+ T cell expansion and function. *J Exp Med* **201**, 139-48 (2005).
46. Tiittanen, M., Westerholm-Ormio, M., Verkasalo, M., Savilahti, E. & Vaarala, O. Infiltration of forkhead box P3-expressing cells in small intestinal mucosa in coeliac disease but not in type 1 diabetes. *Clin Exp Immunol* **152**, 498-507 (2008).
47. Peluso, I. et al. IL-21 counteracts the regulatory T cell-mediated suppression of human CD4+ T lymphocytes. *J Immunol* **178**, 732-9 (2007).
48. Chen, Y., Chou, K., Fuchs, E., Havran, W.L. & Boismenu, R. Protection of the intestinal mucosa by intraepithelial gamma delta T cells. *Proc Natl Acad Sci U S A* **99**, 14338-43 (2002).
49. Bhagat, G. et al. Small intestinal CD8+TCRgammadelta+NKG2A+ intraepithelial lymphocytes have attributes of regulatory cells in patients with celiac disease. *J Clin Invest* **118**, 281-93 (2008).
50. Greco, L. et al. The first large population based twin study of coeliac disease. *Gut* **50**, 624-8 (2002).
51. Sollid, L.M. Coeliac disease: dissecting a complex inflammatory disorder. *Nat Rev Immunol* **2**, 647-55 (2002).
52. Falchuk, Z.M., Rogentine, G.N. & Strober, W. Predominance of histocompatibility antigen HL-A8 in patients with gluten-sensitive enteropathy. *J Clin Invest* **51**, 1602-5 (1972).

53. Keuning, J.J., Pena, A.S., van Leeuwen, A., van Hooff, J.P. & va Rood, J.J. HLA-DW3 associated with coeliac disease. *Lancet* **1**, 506-8 (1976).
54. Alper, C.A., Fleischnick, E., Awdeh, Z., Katz, A.J. & Yunis, E.J. Extended major histocompatibility complex haplotypes in patients with gluten-sensitive enteropathy. *J Clin Invest* **79**, 251-6 (1987).
55. Congia, M. et al. A high frequency of the A30, B18, DR3, DRw52, DQw2 extended haplotype in Sardinian celiac disease patients: further evidence that disease susceptibility is conferred by DQ A1*0501, B1*0201. *Tissue Antigens* **39**, 78-83 (1992).
56. DeMarchi, M. et al. Two HLA-D and DR alleles are associated with coeliac disease. *Tissue Antigens* **14**, 309-16 (1979).
57. Betuel, H. et al. Adult celiac disease associated with HLA-DRw3 and -DRw7. *Tissue Antigens* **15**, 231-8 (1980).
58. Sollid, L.M. et al. Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. *J Exp Med* **169**, 345-50 (1989).
59. Sollid, L.M. Molecular basis of celiac disease. *Annu Rev Immunol* **18**, 53-81 (2000).
60. Karell, K. et al. HLA types in celiac disease patients not carrying the DQA1*05-DQB1*02 (DQ2) heterodimer: results from the European Genetics Cluster on Celiac Disease. *Hum Immunol* **64**, 469-77 (2003).
61. Louka, A.S. et al. HLA in coeliac disease families: a novel test of risk modification by the 'other' haplotype when at least one DQA1*05-DQB1*02 haplotype is carried. *Tissue Antigens* **60**, 147-54 (2002).
62. Vader, W. et al. The HLA-DQ2 gene dose effect in celiac disease is directly related to the magnitude and breadth of gluten-specific T cell responses. *Proc Natl Acad Sci U S A* **100**, 12390-5 (2003).
63. Dubois, P.C. & van Heel, D.A. Translational mini-review series on the immunogenetics of gut disease: immunogenetics of coeliac disease. *Clin Exp Immunol* **153**, 162-73 (2008).
64. Tosi, R. et al. A radioimmunoassay typing study of non-DQw2-associated celiac disease. *Clin Immunol Immunopathol* **39**, 168-72 (1986).
65. Spurkland, A., Sollid, L.M., Polanco, I., Vartdal, F. & Thorsby, E. HLA-DR and -DQ genotypes of celiac disease patients serologically typed to be non-DR3 or non-DR5/7. *Hum Immunol* **35**, 188-92 (1992).

66. Megiorni, F. et al. HLA-DQ and risk gradient for celiac disease. *Hum Immunol* **70**, 55-9 (2009).
67. Petronzelli, F. et al. Genetic contribution of the HLA region to the familial clustering of coeliac disease. *Ann Hum Genet* **61**, 307-17 (1997).
68. Bevan, S. et al. Contribution of the MHC region to the familial risk of coeliac disease. *J Med Genet* **36**, 687-90 (1999).
69. Lewis, C. et al. Celiac disease and human leukocyte antigen genotype: accuracy of diagnosis in self-diagnosed individuals, dosage effect, and sibling risk. *J Pediatr Gastroenterol Nutr* **31**, 22-7 (2000).
70. McManus, R. et al. Association of celiac disease with microsatellite polymorphisms close to the tumor necrosis factor genes. *Hum Immunol* **45**, 24-31 (1996).
71. McManus, R. et al. TNF2, a polymorphism of the tumour necrosis-alpha gene promoter, is a component of the celiac disease major histocompatibility complex haplotype. *Eur J Immunol* **26**, 2113-8 (1996).
72. Polvi, A., Maki, M., Collin, P. & Partanen, J. TNF microsatellite alleles a2 and b3 are not primarily associated with celiac disease in the Finnish population. *Tissue Antigens* **51**, 553-5 (1998).
73. de la Concha, E.G. et al. Celiac disease and TNF promoter polymorphisms. *Hum Immunol* **61**, 513-7 (2000).
74. Garrote, J.A. et al. TNF alpha and LT alpha gene polymorphisms as additional markers of celiac disease susceptibility in a DQ2-positive population. *Immunogenetics* **54**, 551-5 (2002).
75. Fernandez, L. et al. Triplet repeat polymorphism in the transmembrane region of the MICA gene in celiac disease. *Tissue Antigens* **59**, 219-22 (2002).
76. Lopez-Vazquez, A. et al. MHC class I chain related gene A (MICA) modulates the development of coeliac disease in patients with the high risk heterodimer DQA1*0501/DQB1*0201. *Gut* **50**, 336-40 (2002).
77. Bolognesi, E. et al. Additional factor in some HLA DR3/DQ2 haplotypes confers a fourfold increased genetic risk of celiac disease. *Tissue Antigens* **61**, 308-16 (2003).
78. Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).

79. Venter, J.C. et al. The sequence of the human genome. *Science* **291**, 1304-51 (2001).
80. Frazer, K.A., Murray, S.S., Schork, N.J. & Topol, E.J. Human genetic variation and its contribution to complex traits. *Nat Rev Genet* **10**, 241-51 (2009).
81. Lander, E. & Kruglyak, L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* **11**, 241-7 (1995).
82. Altmuller, J., Palmer, L.J., Fischer, G., Scherb, H. & Wjst, M. Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet* **69**, 936-50 (2001).
83. Forabosco, P. et al. Meta-analysis of genome-wide linkage studies in celiac disease. *Hum Hered* **68**, 223-30 (2009).
84. Garner, C.P. et al. Genome-wide linkage analysis of 160 North American families with celiac disease. *Genes Immun* **8**, 108-14 (2007).
85. Greco, L. et al. Genome search in celiac disease. *Am J Hum Genet* **62**, 669-75 (1998).
86. Naluai, A.T. et al. Genome-wide linkage analysis of Scandinavian affected sib-pairs supports presence of susceptibility loci for celiac disease on chromosomes 5 and 11. *Eur J Hum Genet* **9**, 938-44 (2001).
87. Babron, M.C. et al. Meta and pooled analysis of European coeliac disease data. *Eur J Hum Genet* **11**, 828-34 (2003).
88. Ma, Y. et al. A genome-wide search identifies potential new susceptibility loci for Crohn's disease. *Inflamm Bowel Dis* **5**, 271-8 (1999).
89. Rioux, J.D. et al. Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am J Hum Genet* **66**, 1863-70 (2000).
90. Peltekova, V.D. et al. Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nat Genet* **36**, 471-5 (2004).
91. Popat, S. et al. Genome screening of coeliac disease. *J Med Genet* **39**, 328-31 (2002).
92. Van Belzen, M.J. et al. A major non-HLA locus in celiac disease maps to chromosome 19. *Gastroenterology* **125**, 1032-41 (2003).

93. Monsuur, A.J. et al. Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. *Nat Genet* **37**, 1341-4 (2005).
94. Giordano, M. et al. A family-based study does not confirm the association of MYO9B with celiac disease in the Italian population. *Genes Immun* **7**, 606-8 (2006).
95. Hunt, K.A. et al. Lack of association of MYO9B genetic variants with coeliac disease in a British cohort. *Gut* **55**, 969-72 (2006).
96. Nunez, C. et al. No evidence of association of the MYO9B polymorphisms with celiac disease in the Spanish population. *Tissue Antigens* **68**, 489-92 (2006).
97. Abel, M. et al. Adulthood-onset celiac disease is associated with intercellular adhesion molecule-1 (ICAM-1) gene polymorphism. *Hum Immunol* **67**, 612-7 (2006).
98. Cardon, L.R. & Bell, J.I. Association study designs for complex diseases. *Nat Rev Genet* **2**, 91-9 (2001).
99. Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* **22**, 139-44 (1999).
100. Balding, D.J. A tutorial on statistical methods for population association studies. *Nat Rev Genet* **7**, 781-91 (2006).
101. Consortium, I.H. A haplotype map of the human genome. *Nature* **437**, 1299-320 (2005).
102. Frazer, K.A. et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-61 (2007).
103. Wall, J.D. & Pritchard, J.K. Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet* **4**, 587-97 (2003).
104. Chapman, J.M., Cooper, J.D., Todd, J.A. & Clayton, D.G. Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power. *Hum Hered* **56**, 18-31 (2003).
105. Slatkin, M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* **9**, 477-85 (2008).
106. Laird, N.M. & Lange, C. Family-based designs in the age of large-scale gene-association studies. *Nat Rev Genet* **7**, 385-94 (2006).

107. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
108. Djilali-Saiah, I. et al. CTLA-4 gene polymorphism is associated with predisposition to coeliac disease. *Gut* **43**, 187-9 (1998).
109. Popat, S. et al. Variation in the CTLA4/CD28 gene region confers an increased risk of coeliac disease. *Ann Hum Genet* **66**, 125-37 (2002).
110. Jahromi, M.M. & Eisenbarth, G.S. Genetic determinants of type 1 diabetes across populations. *Ann N Y Acad Sci* **1079**, 289-99 (2006).
111. Chistiakov, D.A. & Turakulov, R.I. CTLA-4 and its role in autoimmune thyroid disease. *J Mol Endocrinol* **31**, 21-36 (2003).
112. Haimila, K. et al. The shared CTLA4-ICOS risk locus in celiac disease, IgA deficiency and common variable immunodeficiency. *Genes Immun* **10**, 151-61 (2009).
113. Daha, N.A. et al. Confirmation of STAT4, IL2/IL21, and CTLA4 polymorphisms in rheumatoid arthritis. *Arthritis Rheum* **60**, 1255-60 (2009).
114. Dienz, O. & Rincon, M. The effects of IL-6 on CD4 T cell responses. *Clin Immunol* **130**, 27-33 (2009).
115. Acosta-Rodriguez, E.V., Napolitani, G., Lanzavecchia, A. & Sallusto, F. Interleukins 1beta and 6 but not transforming growth factor-beta are essential for the differentiation of interleukin 17-producing human T helper cells. *Nat Immunol* **8**, 942-9 (2007).
116. Reith, W., LeibundGut-Landmann, S. & Waldburger, J.M. Regulation of MHC class II gene expression by the class II transactivator. *Nat Rev Immunol* **5**, 793-806 (2005).
117. Keklikoglu, N., Koray, M., Kocaelli, H. & Akinci, S. iNOS expression in oral and gastrointestinal tract mucosa. *Dig Dis Sci* **53**, 1437-42 (2008).
118. Warpeha, K.M. et al. Genotyping and functional analysis of a polymorphic (CCTTT)(n) repeat of NOS2A in diabetic retinopathy. *Faseb J* **13**, 1825-32 (1999).
119. Jahromi, M.M., Millward, B.A. & Demaine, A.G. A polymorphism in the promoter region of the gene for interleukin-6 is associated with susceptibility to type 1 diabetes mellitus. *J Interferon Cytokine Res* **20**, 885-8 (2000).
120. Rasmussen, H.B., Kelly, M.A. & Clausen, J. Genetic susceptibility to multiple sclerosis: detection of polymorphic nucleotides and an intron in the 3'

untranslated region of the major histocompatibility complex class II transactivator gene. *Hum Immunol* **62**, 371-7 (2001).

121. Castellanos-Rubio, A. et al. Combined functional and positional gene information for the identification of susceptibility variants in celiac disease. *Gastroenterology* **134**, 738-46 (2008).
122. Hunt, K.A., Franke, L., Deloukas, P., Wijmenga, C. & van Heel, D.A. No evidence in a large UK collection for celiac disease risk variants reported by a Spanish study. *Gastroenterology* **134**, 1629-30; author reply 1630-1 (2008).
123. Duerr, R.H. et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* **314**, 1461-3 (2006).
124. Consortium, W.T.C.C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-78 (2007).
125. McCarthy, M.I. et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* **9**, 356-69 (2008).
126. van Heel, D.A. et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* **39**, 827-9 (2007).
127. Smyth, D.J. et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. *N Engl J Med* **359**, 2767-77 (2008).
128. Todd, J.A. et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* **39**, 857-64 (2007).
129. Liu, Y. et al. A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. *PLoS Genet* **4**, e1000041 (2008).
130. Zhernakova, A. et al. Novel association in chromosome 4q27 region with rheumatoid arthritis and confirmation of type 1 diabetes point to a general risk locus for autoimmune diseases. *Am J Hum Genet* **81**, 1284-8 (2007).
131. Glas, J. et al. Novel genetic risk markers for ulcerative colitis in the IL2/IL21 region are in epistasis with IL23R and suggest a common genetic background for ulcerative colitis and celiac disease. *Am J Gastroenterol* **104**, 1737-44 (2009).
132. Hunt, K.A. et al. Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet* **40**, 395-402 (2008).

133. Garner, C.P. et al. Replication of celiac disease UK genome-wide association study results in a US population. *Hum Mol Genet* **18**, 4219-25 (2009).
134. Koskinen, L.L. et al. Association study of the IL18RAP locus in three European populations with coeliac disease. *Hum Mol Genet* **18**, 1148-55 (2009).
135. Romanos, J. et al. Six new coeliac disease loci replicated in an Italian population confirm association with coeliac disease. *J Med Genet* **46**, 60-3 (2009).
136. Hollinger, S. & Hepler, J.R. Cellular regulation of RGS proteins: modulators and integrators of G protein signaling. *Pharmacol Rev* **54**, 527-59 (2002).
137. Arend, W.P., Palmer, G. & Gabay, C. IL-1, IL-18, and IL-33 families of cytokines. *Immunol Rev* **223**, 20-38 (2008).
138. Okazawa, A. et al. Human intestinal epithelial cell-derived interleukin (IL)-18, along with IL-2, IL-7 and IL-15, is a potent synergistic factor for the proliferation of intraepithelial lymphocytes. *Clin Exp Immunol* **136**, 269-76 (2004).
139. Zhernakova, A. et al. Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring CARD9 and IL18RAP. *Am J Hum Genet* **82**, 1202-10 (2008).
140. Charo, I.F. & Ransohoff, R.M. The many roles of chemokines and chemokine receptors in inflammation. *N Engl J Med* **354**, 610-21 (2006).
141. Papadakis, K.A. Chemokines in inflammatory bowel disease. *Curr Allergy Asthma Rep* **4**, 83-9 (2004).
142. Mamtani, M. et al. CCL3L1 gene-containing segmental duplications and polymorphisms in CCR5 affect risk of systemic lupus erythaematosus. *Ann Rheum Dis* **67**, 1076-83 (2008).
143. McKinney, C. et al. Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. *Ann Rheum Dis* **67**, 409-13 (2008).
144. Hunter, C.A. New IL-12-family members: IL-23 and IL-27, cytokines with divergent functions. *Nat Rev Immunol* **5**, 521-31 (2005).
145. Coenen, M.J. et al. Common and different genetic background for rheumatoid arthritis and coeliac disease. *Hum Mol Genet* **18**, 4195-203 (2009).

146. Heasman, S.J. & Ridley, A.J. Mammalian Rho GTPases: new insights into their functions from in vivo studies. *Nat Rev Mol Cell Biol* **9**, 690-701 (2008).
147. Velazquez, L. et al. Cytokine signaling and hematopoietic homeostasis are disrupted in Lnk-deficient mice. *J Exp Med* **195**, 1599-611 (2002).
148. Takaki, S. et al. Impaired lymphopoiesis and altered B cell subpopulations in mice overexpressing Lnk adaptor protein. *J Immunol* **170**, 703-10 (2003).
149. Zhernakova, A., van Diemen, C.C. & Wijmenga, C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat Rev Genet* **10**, 43-55 (2009).
150. Trynka, G. et al. Coeliac disease-associated risk variants in TNFAIP3 and REL implicate altered NF-kappaB signalling. *Gut* **58**, 1078-83 (2009).
151. Plenge, R.M. et al. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* **39**, 1477-82 (2007).
152. Thomson, W. et al. Rheumatoid arthritis association at 6q23. *Nat Genet* **39**, 1431-3 (2007).
153. Fung, E.Y. et al. Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus. *Genes Immun* **10**, 188-91 (2009).
154. Graham, R.R. et al. Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat Genet* **40**, 1059-61 (2008).
155. Gregersen, P.K. et al. REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nat Genet* **41**, 820-3 (2009).
156. Hayden, M.S., West, A.P. & Ghosh, S. NF-kappaB and the immune response. *Oncogene* **25**, 6758-80 (2006).
157. Romanos, J. et al. Analysis of HLA and non-HLA alleles can identify individuals at high risk for celiac disease. *Gastroenterology* **137**, 834-40, 840 e1-3 (2009).
158. Pabst, O., Zweigerdt, R. & Arnold, H.H. Targeted disruption of the homeobox transcription factor Nkx2-3 in mice results in postnatal lethality and abnormal development of small intestine and spleen. *Development* **126**, 2215-25 (1999).
159. Levine, B. & Deretic, V. Unveiling the roles of autophagy in innate and adaptive immunity. *Nat Rev Immunol* **7**, 767-77 (2007).

160. Bettelli, E., Korn, T. & Kuchroo, V.K. Th17: the third member of the effector T cell trilogy. *Curr Opin Immunol* **19**, 652-7 (2007).
161. Hakonarson, H. et al. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* **448**, 591-4 (2007).
162. Hafler, D.A. et al. Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* **357**, 851-62 (2007).
163. Kagnoff, M.F. Celiac disease: pathogenesis of a model immunogenetic disease. *J Clin Invest* **117**, 41-9 (2007).
164. Siegel, M. et al. Extracellular transglutaminase 2 is catalytically inactive, but is transiently activated upon tissue injury. *PLoS One* **3**, e1861 (2008).
165. Fesus, L. & Piacentini, M. Transglutaminase 2: an enigmatic enzyme with diverse functions. *Trends Biochem Sci* **27**, 534-9 (2002).
166. Molberg, O. et al. Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. *Nat Med* **4**, 713-7 (1998).
167. Sjostrom, H. et al. Identification of a gliadin T-cell epitope in coeliac disease: general importance of gliadin deamidation for intestinal T-cell recognition. *Scand J Immunol* **48**, 111-5 (1998).
168. Godkin, A. et al. Use of eluted peptide sequence data to identify the binding characteristics of peptides to the insulin-dependent diabetes susceptibility allele HLA-DQ8 (DQ 3.2). *Int Immunol* **9**, 905-11 (1997).
169. Johansen, B.H., Vartdal, F., Eriksen, J.A., Thorsby, E. & Sollid, L.M. Identification of a putative motif for binding of peptides to HLA-DQ2. *Int Immunol* **8**, 177-82 (1996).
170. van de Wal, Y., Kooy, Y.M., Drijfhout, J.W., Amons, R. & Koning, F. Peptide binding characteristics of the coeliac disease-associated DQ(alpha1*0501, beta1*0201) molecule. *Immunogenetics* **44**, 246-53 (1996).
171. Qiao, S.W., Sollid, L.M. & Blumberg, R.S. Antigen presentation in celiac disease. *Curr Opin Immunol* **21**, 111-7 (2009).
172. Camarca, A. et al. Intestinal T cell responses to gluten peptides are largely heterogeneous: implications for a peptide-based therapy in celiac disease. *J Immunol* **182**, 4158-66 (2009).

173. Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. *Arch Dis Child* **65**, 909-11 (1990).
174. Miller, S.A., Dykes, D.D. & Polesky, H.F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16**, 1215 (1988).
175. Kimura, A., Sasazuki, T. Eleventh International Histocompatibility Workshop reference protocol for the HLA DNA-typing technique. in *HLA 1991. Proceedings of the Eleventh International Histocompatibility Workshop and Conference*, Vol. 1 (ed. Tsuji, K., Aizawa, M., Sasazuki, T.) 397-419 (Oxford University press, Oxford, 1992).
176. Livak, K.J., Flood, S.J., Marmaro, J., Giusti, W. & Deetz, K. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl* **4**, 357-62 (1995).
177. Weir, B.S. *Genetic data analysis*, (Sinauer Associates, Inc. Sunderland, Massachusetts, 1990).
178. Lewontin, R.C. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics* **49**, 49-67 (1964).
179. Schneider, S., J. M. Kueffer, D. Roessli and L. Excoffier. *Arlequin ver. 2.0: A Software for Population Genetic Data Analysis.*, (University of Geneva. Switzerland, 2000).
180. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-5 (2005).
181. Spielman, R.S., McGinnis, R.E. & Ewens, W.J. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* **52**, 506-16 (1993).
182. Breunis, W.B. et al. Polymorphisms in chemokine receptor genes and susceptibility to Kawasaki disease. *Clin Exp Immunol* **150**, 83-90 (2007).
183. Reijmerink, N.E. et al. Association of IL1RL1, IL18R1, and IL18RAP gene cluster polymorphisms with asthma and atopy. *J Allergy Clin Immunol* **122**, 651-4 e8 (2008).
184. Alderton, W.K., Cooper, C.E. & Knowles, R.G. Nitric oxide synthases: structure, function and inhibition. *Biochem J* **357**, 593-615 (2001).

185. Murray, I.A., Daniels, I., Coupland, K., Smith, J.A. & Long, R.G. Increased activity and expression of iNOS in human duodenal enterocytes from patients with celiac disease. *Am J Physiol Gastrointest Liver Physiol* **283**, G319-26 (2002).
186. De Stefano, D. et al. The role of NF-kappaB, IRF-1, and STAT-1alpha transcription factors in the iNOS gene induction by gliadin and IFN-gamma in RAW 264.7 macrophages. *J Mol Med* **84**, 65-74 (2006).
187. de Vera, M.E. et al. Transcriptional regulation of human inducible nitric oxide synthase (NOS2) gene by cytokines: initial analysis of the human NOS2 promoter. *Proc Natl Acad Sci U S A* **93**, 1054-9 (1996).
188. Kleinert, H., Schwarz, P.M. & Forstermann, U. Regulation of the expression of inducible nitric oxide synthase. *Biol Chem* **384**, 1343-64 (2003).
189. Gonzalez-Gay, M.A. et al. Inducible but not endothelial nitric oxide synthase polymorphism is associated with susceptibility to rheumatoid arthritis in northwest Spain. *Rheumatology (Oxford)* **43**, 1182-85 (2004).
190. Oates, J.C. et al. Nitric oxide synthase 2 promoter polymorphisms and systemic lupus erythematosus in african-americans. *J Rheumatol* **30**, 60-67 (2003).
191. Rueda, B. et al. Polymorphism of the inducible nitric oxide synthase gene in celiac disease. *Hum Immunol* **63**, 1062-65 (2002).
192. Kroncke, K.D., Fehsel, K. & Kolb-Bachofen, V. Inducible nitric oxide synthase in human diseases. *Clin Exp Immunol* **113**, 147-56 (1998).
193. Sato, T. et al. Dual functions of Runx proteins for reactivating CD8 and silencing CD4 at the commitment process into CD8 thymocytes. *Immunity* **22**, 317-28 (2005).
194. Woolf, E. et al. Runx3 and Runx1 are required for CD8 T cell development during thymopoiesis. *Proc Natl Acad Sci U S A* **100**, 7731-6 (2003).
195. Forsberg, G. et al. Presence of bacteria and innate immunity of intestinal epithelium in childhood celiac disease. *Am J Gastroenterol* **99**, 894-904 (2004).
196. Stene, L.C. et al. Rotavirus infection frequency and risk of celiac disease autoimmunity in early childhood: a longitudinal study. *Am J Gastroenterol* **101**, 2333-40 (2006).

197. Singal, D.P. & Blajchman, M.A. Histocompatibility (HL-A) antigens, lymphocytotoxic antibodies and tissue antibodies in patients with diabetes mellitus. *Diabetes* **22**, 429-32 (1973).
198. Hill, A.V. et al. Common west African HLA antigens are associated with protection from severe malaria. *Nature* **352**, 595-600 (1991).
199. Corder, E.H. et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921-3 (1993).
200. Ogura, Y. et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**, 603-6 (2001).
201. Goring, H.H., Terwilliger, J.D. & Blangero, J. Large upward bias in estimation of locus-specific effects from genomewide scans. *Am J Hum Genet* **69**, 1357-69 (2001).
202. Zollner, S. & Pritchard, J.K. Overcoming the winner's curse: estimating penetrance parameters from case-control data. *Am J Hum Genet* **80**, 605-15 (2007).
203. Kavvoura, F.K. & Ioannidis, J.P. Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls. *Hum Genet* **123**, 1-14 (2008).
204. Novembre, J. et al. Genes mirror geography within Europe. *Nature* **456**, 98-101 (2008).
205. Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S. & Hirschhorn, J.N. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* **33**, 177-82 (2003).
206. Fernando, M.M. et al. Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS Genet* **4**, e1000024 (2008).
207. Ueda, H. et al. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* **423**, 506-11 (2003).
208. Vang, T., Miletic, A.V., Bottini, N. & Mustelin, T. Protein tyrosine phosphatase PTPN22 in human autoimmunity. *Autoimmunity* **40**, 453-61 (2007).
209. Vaarala, O., Atkinson, M.A. & Neu, J. The "perfect storm" for type 1 diabetes: the complex interplay between intestinal microbiota, gut permeability, and mucosal immunity. *Diabetes* **57**, 2555-62 (2008).

210. Yang, A. et al. Inflammatory bowel disease in patients with celiac disease. *Inflamm Bowel Dis* **11**, 528-32 (2005).
211. Mahmud, F.H. et al. Celiac disease in type 1 diabetes mellitus in a North American community: prevalence, serologic screening, and clinical features. *Mayo Clin Proc* **80**, 1429-34 (2005).
212. Green, P.H. et al. Mechanisms underlying celiac disease and its neurologic manifestations. *Cell Mol Life Sci* **62**, 791-9 (2005).
213. Briani, C. et al. Neurological complications of celiac disease and autoimmune mechanisms: a prospective study. *J Neuroimmunol* **195**, 171-5 (2008).
214. Ferro, M.T. et al. A case of multiple sclerosis with atypical onset associated with autoimmune hepatitis and silent coeliac disease. *Neurol Sci* **29**, 29-31 (2008).
215. Yamout, B., Usta, J., Itani, S. & Yaghi, S. Celiac disease, Behcet, and idiopathic thrombocytopenic purpura in siblings of a patient with multiple sclerosis. *Mult Scler* **15**, 1368-71 (2009).
216. Hampe, J. et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet* **39**, 207-11 (2007).
217. Einarsdottir, E. et al. IL23R in the Swedish, Finnish, Hungarian and Italian populations: association with IBD and psoriasis, and linkage to celiac disease. *BMC Med Genet* **10**, 8 (2009).
218. Weersma, R.K. et al. ATG16L1 and IL23R are associated with inflammatory bowel diseases but not with celiac disease in the Netherlands. *Am J Gastroenterol* **103**, 621-7 (2008).
219. Cargill, M. et al. A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes. *Am J Hum Genet* **80**, 273-90 (2007).
220. Brown, M.A. Breakthroughs in genetic studies of ankylosing spondylitis. *Rheumatology (Oxford)* **47**, 132-7 (2008).
221. Moore, J.H. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Hum Hered* **56**, 73-82 (2003).
222. Carlborg, O. & Haley, C.S. Epistasis: too often neglected in complex trait studies? *Nat Rev Genet* **5**, 618-25 (2004).

223. Gregersen, J.W. et al. Functional epistasis on a common MHC haplotype associated with multiple sclerosis. *Nature* **443**, 574-7 (2006).
224. Fisher, R.A. The correlation between relatives on the supposition of Mendelian inheritance. *Proc R Soc Edinburgh* **52**, 399-433 (1918).
225. Sartoris, S. et al. Analysis of CIITA encoding AIR-1 gene promoters in insulin-dependent diabetes mellitus and rheumatoid arthritis patients from the northeast of Italy: absence of sequence variability. *Hum Immunol* **61**, 599-604 (2000).
226. Martinez, A. et al. Chromosomal region 16p13: further evidence of increased predisposition to immune diseases. *Ann Rheum Dis* **69**, 309-11 (2010).
227. Baranzini, S.E. The genetics of autoimmune diseases: a networked perspective. *Curr Opin Immunol* **21**, 596-605 (2009).
228. Diosdado, B. et al. A microarray screen for novel candidate genes in coeliac disease pathogenesis. *Gut* **53**, 944-51 (2004).
229. Juuti-Uusitalo, K. et al. cDNA microarray analysis of gene expression in coeliac disease jejunal biopsy samples. *J Autoimmun* **22**, 249-65 (2004).
230. Bracken, S., Byrne, G., Kelly, J., Jackson, J. & Feighery, C. Altered gene expression in highly purified enterocytes from patients with active coeliac disease. *BMC Genomics* **9**, 377 (2008).
231. Manavalan, J.S. et al. Serum cytokine elevations in celiac disease: association with disease presentation. *Hum Immunol* **71**, 50-7 (2010).
232. Romaldini, C.C., Barbieri, D., Okay, T.S., Raiz, R., Jr. & Cancado, E.L. Serum soluble interleukin-2 receptor, interleukin-6, and tumor necrosis factor-alpha levels in children with celiac disease: response to treatment. *J Pediatr Gastroenterol Nutr* **35**, 513-7 (2002).
233. Daniels, I., Cavill, D., Murray, I.A. & Long, R.G. Elevated expression of iNOS mRNA and protein in coeliac disease. *Clin Chim Acta* **356**, 134-42 (2005).
234. Fishman, D. et al. The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. *J Clin Invest* **102**, 1369-76 (1998).
235. Terry, C.F., Loukaci, V. & Green, F.R. Cooperative influence of genetic polymorphisms on interleukin 6 transcriptional regulation. *J Biol Chem* **275**, 18138-44 (2000).

236. Ray, P., Ghosh, S.K., Zhang, D.H. & Ray, A. Repression of interleukin-6 gene expression by 17 beta-estradiol: inhibition of the DNA-binding activity of the transcription factors NF-IL6 and NF-kappa B by the estrogen receptor. *FEBS Lett* **409**, 79-85 (1997).

237. Kristiansen, O.P. et al. Association of a functional 17beta-estradiol sensitive IL6-174G/C promoter polymorphism with early-onset type 1 diabetes in females. *Hum Mol Genet* **12**, 1101-10 (2003).

238. Woolley, N., Mustalahti, K., Maki, M. & Partanen, J. Cytokine gene polymorphisms and genetic association with coeliac disease in the Finnish population. *Scand J Immunol* **61**, 51-6 (2005).

239. Fish, E.N. The X-files in immunity: sex-based differences predispose immune responses. *Nat Rev Immunol* **8**, 737-44 (2008).

240. Qi, L., Rifai, N. & Hu, F.B. Interleukin-6 receptor gene variations, plasma interleukin-6 levels, and type 2 diabetes in U.S. Women. *Diabetes* **56**, 3075-81 (2007).

241. Reich, D. et al. Admixture mapping of an allele affecting interleukin 6 soluble receptor and interleukin 6 levels. *Am J Hum Genet* **80**, 716-26 (2007).

242. Mullberg, J. et al. The soluble human IL-6 receptor. Mutational characterization of the proteolytic cleavage site. *J Immunol* **152**, 4958-68 (1994).

243. Morris, B.J. et al. Association of a functional inducible nitric oxide synthase promoter variant with complications in type 2 diabetes. *J Mol Med* **80**, 96-104 (2002).

244. Manna, I. et al. Preliminary evidences of a NOS2A protective effect from relapsing-remitting multiple sclerosis. *J Neurol Sci* **264**, 112-7 (2008).

245. Burgner, D. et al. Haplotypic relationship between SNP and microsatellite markers at the NOS2A locus in two populations. *Genes Immun* **4**, 506-14 (2003).

246. Harris, K.M., Fasano, A. & Mann, D.L. Cutting edge: IL-1 controls the IL-23 response induced by gliadin, the etiologic agent in celiac disease. *J Immunol* **181**, 4457-60 (2008).

247. Przemioslo, R.T., Kontakou, M., Nobili, V. & Ciclitira, P.J. Raised pro-inflammatory cytokines interleukin 6 and tumour necrosis factor alpha in

coeliac disease mucosa detected by immunohistochemistry. *Gut* **35**, 1398-403 (1994).

248. Lahat, N. et al. Cytokine profile in coeliac disease. *Scand J Immunol* **49**, 441-6 (1999).
249. Fornari, M.C. et al. Pre- and post-treatment serum levels of cytokines IL-1beta, IL-6, and IL-1 receptor antagonist in celiac disease. Are they related to the associated osteopenia? *Am J Gastroenterol* **93**, 413-8 (1998).
250. Puren, A.J., Razeghi, P., Fantuzzi, G. & Dinarello, C.A. Interleukin-18 enhances lipopolysaccharide-induced interferon-gamma production in human whole blood cultures. *J Infect Dis* **178**, 1830-4 (1998).
251. Kishimoto, T., Akira, S. & Taga, T. Interleukin-6 and its receptor: a paradigm for cytokines. *Science* **258**, 593-7 (1992).
252. Izcue, A., Coombes, J.L. & Powrie, F. Regulatory lymphocytes and intestinal inflammation. *Annu Rev Immunol* **27**, 313-38 (2009).
253. Bettelli, E. et al. Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells. *Nature* **441**, 235-8 (2006).
254. Manel, N., Unutmaz, D. & Littman, D.R. The differentiation of human T(H)-17 cells requires transforming growth factor-beta and induction of the nuclear receptor RORgammat. *Nat Immunol* **9**, 641-9 (2008).
255. Yang, L. et al. IL-21 and TGF-beta are required for differentiation of human T(H)17 cells. *Nature* **454**, 350-2 (2008).
256. Zhou, L. et al. IL-6 programs T(H)-17 cell differentiation by promoting sequential engagement of the IL-21 and IL-23 pathways. *Nat Immunol* **8**, 967-74 (2007).
257. Castellanos-Rubio, A. et al. TH17 (and TH1) signatures of intestinal biopsies of CD patients in response to gliadin. *Autoimmunity* **42**, 69-73 (2009).
258. Korn, T. et al. IL-21 initiates an alternative pathway to induce proinflammatory T(H)17 cells. *Nature* **448**, 484-7 (2007).
259. Weaver, C.T., Harrington, L.E., Mangan, P.R., Gavrieli, M. & Murphy, K.M. Th17: an effector CD4 T cell lineage with regulatory T cell ties. *Immunity* **24**, 677-88 (2006).
260. Gutcher, I., Urich, E., Wolter, K., Prinz, M. & Becher, B. Interleukin 18-independent engagement of interleukin 18 receptor-alpha is required for autoimmune inflammation. *Nat Immunol* **7**, 946-53 (2006).

261. Acosta-Rodriguez, E.V. et al. Surface phenotype and antigenic specificity of human interleukin 17-producing T helper memory cells. *Nat Immunol* **8**, 639-46 (2007).
262. Sato, W., Aranami, T. & Yamamura, T. Cutting edge: Human Th17 cells are identified as bearing CCR2+CCR5- phenotype. *J Immunol* **178**, 7525-9 (2007).
263. Strengell, M. et al. IL-21 in synergy with IL-15 or IL-18 enhances IFN-gamma production in human NK and T cells. *J Immunol* **170**, 5464-9 (2003).
264. Yoshimoto, T. et al. IL-12 up-regulates IL-18 receptor expression on T cells, Th1 cells, and B cells: synergism with IL-18 for IFN-gamma production. *J Immunol* **161**, 3400-7 (1998).
265. Zabel, B.A. et al. Human G protein-coupled receptor GPR-9-6/CC chemokine receptor 9 is selectively expressed on intestinal homing T lymphocytes, mucosal lymphocytes, and thymocytes and is required for thymus-expressed chemokine-mediated chemotaxis. *J Exp Med* **190**, 1241-56 (1999).
266. Uehara, S., Grinberg, A., Farber, J.M. & Love, P.E. A role for CCR9 in T lymphocyte development and migration. *J Immunol* **168**, 2811-9 (2002).
267. Olaussen, R.W. et al. Reduced chemokine receptor 9 on intraepithelial lymphocytes in celiac disease suggests persistent epithelial activation. *Gastroenterology* **132**, 2371-82 (2007).
268. Jabri, B. & Sollid, L.M. Tissue-mediated control of immunopathology in coeliac disease. *Nat Rev Immunol* **9**, 858-70 (2009).